

A Nesterov-Accelerated Primal–Dual Splitting Algorithm for Convex Nonsmooth Optimization

Laurent Condat*, Abdurakhmon Sadiev, and Peter Richtárik

King Abdullah University of Science and Technology (KAUST),
Thuwal, Kingdom of Saudi Arabia

April 10, 2026

Abstract

We investigate the integration of Nesterov-type acceleration into primal–dual methods for structured convex optimization. While proximal splitting algorithms efficiently handle composite problems of the form $\min_x f(x) + g(x) + h(Kx)$, accelerating their convergence with respect to the smooth term f is notoriously challenging due to the rotational dynamics in the primal–dual space. In this paper, we overcome this barrier by proposing the Accelerated Proximal Alternating Predictor–Corrector algorithm (APAPC), focusing on the setting where $g(x) = \frac{\mu_g}{2} \|x\|^2$. Our analysis reveals that Nesterov momentum can be seamlessly integrated into a primal–dual forward–backward scheme by exploiting the strong convexity of the dual problem to stabilize the accelerated primal updates. Using a unified Lyapunov framework, we establish optimal $\mathcal{O}(1/t^2)$ sublinear convergence rates, as well as accelerated linear convergence when $\mu_g > 0$, across three regimes of dual strong convexity: (i) when h is smooth, (ii) when the linear operator K^* is bounded below, and (iii) for linearly constrained optimization. Furthermore, leveraging recent results on accelerated gradient descent, we characterize the weak convergence of the primal–dual iterates to a saddle-point solution.

1 Introduction

Let \mathcal{X} and \mathcal{U} be real Hilbert spaces. We consider structured convex optimization problems of the form

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \Psi(x) := f(x) + g(x) + h(Kx), \quad (1)$$

where $f : \mathcal{X} \rightarrow \mathbb{R}$ is convex and differentiable, $g : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ and $h : \mathcal{U} \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper lower semicontinuous convex functions, and $K : \mathcal{X} \rightarrow \mathcal{U}$ is a nonzero bounded linear operator [Bauschke and Combettes, 2017]. We assume that f is L_f -smooth for some $L_f > 0$, that is, its gradient ∇f is L_f -Lipschitz continuous on \mathcal{X} .

Associated with (1), we consider the dual problem [Bauschke and Combettes, 2017, Section 15.3]

$$\underset{u \in \mathcal{U}}{\text{minimize}} \quad (f + g)^*(-K^*u) + h^*(u), \quad (2)$$

*Corresponding author. Contact: see <https://lcondat.github.io/>

where ϕ^* denotes the convex conjugate of a function ϕ [Bauschke and Combettes, 2017, Chapter 13], and K^* is the adjoint of K . For any convex function ϕ , we denote by $\mu_\phi \geq 0$ a constant such that ϕ is μ_ϕ -strongly convex, that is, $\phi - \frac{\mu_\phi}{2} \|\cdot\|^2$ is convex.

Problems of the form (1) arise in signal and image processing, inverse problems, control, and machine learning, where g and h encode nonsmooth regularization terms and constraints [Palomar and Eldar, 2009, Sra et al., 2011, Bach et al., 2012, Cevher et al., 2014, Polson et al., 2015, Glowinski et al., 2016, Chambolle and Pock, 2016a, Stathopoulos et al., 2016, Bottou et al., 2018].

We recall that, for a proper lower semicontinuous convex function ϕ on \mathcal{X} , its proximity operator is the mapping $\text{prox}_\phi : x \in \mathcal{X} \mapsto \arg \min_{y \in \mathcal{X}} \left\{ \phi(y) + \frac{1}{2} \|y - x\|^2 \right\}$, which is well defined and single-valued [Parikh and Boyd, 2014]. We study fully-split primal–dual proximal algorithms; that is, methods that solve the primal and dual problems (1) and (2) simultaneously by iteratively evaluating K , K^* , ∇f , and the proximity operators of g and h (or equivalently h^* via the Moreau identity $x = \text{prox}_{h/\tau}(x) + \frac{1}{\tau} \text{prox}_{\tau h^*}(\tau x)$). [Bauschke and Combettes, 2017, Theorem 14.3(ii)].

For the primal and dual problems (1)–(2) to be well posed, we assume throughout that there exists $(x^*, u^*) \in \mathcal{X} \times \mathcal{U}$ satisfying the optimality conditions

$$\begin{cases} 0 \in \nabla f(x^*) + \partial g(x^*) + K^* u^*, \\ 0 \in \partial h^*(u^*) - K x^*. \end{cases} \quad (3)$$

Under these conditions, x^* is a solution of (1) and u^* is a solution of (2). Moreover, if a suitable constraint qualification holds, then (3) is not only sufficient but also necessary for (x^*, u^*) to be a primal–dual solution [Bauschke and Combettes, 2017, Section 27.1].

This work investigates how Nesterov-type acceleration [Nesterov, 2013] can be effectively incorporated into primal–dual methods for this class of problems. Section 2 reviews Nesterov’s acceleration of proximal gradient descent for the minimization of $f + g$. Section 3 then introduces our new algorithm, which achieves the same accelerated dependence on $f + g$ while accommodating the composite term $h \circ K$. In the remainder of this section, we review the state of the art and detail our contributions.

1.1 Related Work

Algorithms that solve nonsmooth optimization problems involving several functions through iterative evaluations of gradients and proximity operators are known as *proximal splitting algorithms* [Combettes and Pesquet, 2010, Boş et al., 2014, Parikh and Boyd, 2014, Beck, 2017, Combettes and Pesquet, 2021]. For Problem (1), which involves the composite term $h \circ K$, several primal–dual algorithms have been proposed. We refer to the surveys Condat et al. [2023], Combettes [2024] and to the recent developments in Salim et al. [2022b], Condat et al. [2022]. We briefly review the most relevant methods, focusing on linear convergence results.

Proximal Gradient Descent. A prominent method for the simple problem of minimizing $f + g$ is *Proximal Gradient Descent* (PGD), which iterates $x^{t+1} := \text{prox}_{\gamma g}(x^t - \gamma \nabla f(x^t))$, where $t \geq 0$ is the iteration index and $\gamma > 0$ the stepsize. It is known that if $\gamma \in (0, \frac{2}{L_f})$, PGD achieves a convergence rate $\mathcal{O}(1/t)$ [Bubeck, 2015, Taylor et al., 2018]. If, in addition, $\mu_f > 0$ or $\mu_g > 0$, PGD converges linearly to the unique solution x^* , with $\|x^{t+1} - x^*\| \leq \frac{\max(1 - \gamma \mu_f, \gamma L_f - 1)}{1 + \gamma \mu_g} \|x^t - x^*\|$ [Polyak, 1963, Theorem 6], [Condat and Richtárik, 2023, Lemma 1]. These rates are not accelerated and can be improved using Nesterov-type acceleration [Nesterov, 2013], as discussed in Section 2.

Primal–Dual Hybrid Gradient and Condat–Vũ algorithms. A key method is the *Primal–Dual Hybrid Gradient* algorithm (PDHG) for minimizing $g + h \circ K$. A modified version of PDHG that requires knowledge of μ_g converges linearly when $\mu_g > 0$ and h is L_h -smooth. When the stepsizes are chosen optimally, its complexity, as a number of iterations to reach a given accuracy, is accelerated and scales as $\sqrt{\|K\|^2 L_h / \mu_g}$ [Chambolle and Pock, 2011].¹ The unmodified PDHG has been shown to converge linearly under the same assumptions [O’Connor and Vandenberghe, 2020],[Condat and Richtárik, 2023, Theorem 7].

To extend PDHG to Problem (1) with $f \neq 0$, the *Condat–Vũ* algorithm (CV) was proposed independently by Condat [2013] and Vũ [2013]. CV reduces to PDHG when $f = 0$. With stepsizes $\gamma > 0$ and $\tau > 0$, it iterates

$$\begin{aligned} x^{t+1} &:= \text{prox}_{\gamma g}(x^t - \gamma \nabla f(x^t) - \gamma K^* u^t), \\ u^{t+1} &:= \text{prox}_{\tau h^*}(u^t + \tau K(2x^{t+1} - x^t)) \end{aligned}$$

(up to a symmetric form obtained by swapping the order of the updates). Linear convergence of CV has been studied in Chambolle and Pock [2016b], Dirren et al. [2025].

Proximal Alternating Predictor–Corrector and three-operator splitting algorithms. Another method for Problem (1) with $g = 0$ is the *Proximal Alternating Predictor–Corrector* algorithm (PAPC) [Drori et al., 2015], also proposed independently as PDFP²O [Chen et al., 2013] and in Loris and Verhoeven [2011] for regularized least-squares problems. With stepsizes $\gamma > 0$ and $\tau > 0$, PAPC iterates

$$\begin{aligned} \hat{x}^t &:= x^t - \gamma \nabla f(x^t) - \gamma K^* u^t, \\ u^{t+1} &:= \text{prox}_{\tau h^*}(u^t + \tau K \hat{x}^t), \\ x^{t+1} &:= x^t - \gamma \nabla f(x^t) - \gamma K^* u^{t+1}. \end{aligned} \tag{4}$$

The method converges whenever $\gamma \in (0, \frac{2}{L_f})$ and $\gamma \tau \|K\|^2 \leq 1$, which provides a larger admissible stepsize range than CV [Condat et al., 2023]. Both CV and PAPC can be interpreted as preconditioned forward–backward iterations in the product space $\mathcal{X} \times \mathcal{U}$ [Combettes et al., 2014], differing only in the choice of preconditioner.

When $K = I$ is the identity, the *Three-Operator Splitting* algorithm (TOS) is well suited to minimize $f + g + h$ [Davis and Yin, 2017], and its linear convergence has been studied in Lee et al. [2025]. More generally, two “twin” algorithms have been proposed for Problem (1), recovering PDHG, PAPC and TOS as special cases: the *Primal–Dual Three-Operator Splitting* algorithm (PD3O) [Yan, 2018] and the *Primal–Dual Davis–Yin* algorithm (PDDY) [Salim et al., 2022b]. Both methods converge under the same stepsize conditions as PAPC. Their linear convergence was established in [Condat et al., 2022], and refined rates for PDDY (and hence PDHG and PAPC) were derived in Condat and Richtárik [2023].

When $\mu_f > 0$ or $\mu_g > 0$ and h is L_h -smooth, the complexity of CV, PD3O and PDDY, optimized over the stepsizes, scales as $\frac{L_f}{\mu_f + \mu_g} + \sqrt{\frac{\|K\|^2 L_h}{\mu_f + \mu_g}}$. The second term exhibits accelerated dependence on the primal–dual condition number, whereas the first term matches the complexity of PGD.

¹If h is L_h -smooth, $h \circ K$ is $\|K\|^2 L_h$ -smooth, so $\|K\|^2 L_h / \mu_g$ can be interpreted as the condition number of $g + h \circ K$. Moreover, h^* is $1/L_h$ -strongly convex and $g^* \circ -K^*$ is $\|K\|^2 / \mu_g$ -smooth, so the same quantity also characterizes the condition number of $g^* \circ -K^* + h^*$. Thus, PDHG can be regarded as accelerated in the sense that its complexity depends on the square root of this primal–dual condition number.

This behavior is expected, since these algorithms treat f via a standard gradient descent step $x^t - \gamma \nabla f(x^t)$.

Acceleration of primal–dual algorithms. A natural question is therefore whether these primal–dual schemes can be accelerated so as to achieve improved dependence on L_f . Acceleration of primal–dual algorithms is notoriously challenging. They have a rotational dynamics in the primal–dual space, and momentum-based acceleration applied naively will exacerbate this effect and cause the method to diverge. Accelerated variants of **CV** have been proposed, initially for $g = 0$ [Chen et al., 2014] and later in the general case [Zhao et al., 2019, Zhao, 2022], but without linear convergence guarantees. Recently, a comprehensive analysis of the *Accelerated Condat–Vũ* algorithm (**ACV**) was presented [Driggs et al., 2024]. Notably, it achieves accelerated linear convergence when $\mu_g > 0$ and h is L_h -smooth, with complexity scaling as $\sqrt{\frac{L_f}{\mu_g}} + \sqrt{\frac{\|K\|^2 L_h}{\mu_g}}$.

1.2 Contributions

In this work, we study the acceleration of **PAPC**, which has a primal–dual forward–backward structure similar to **CV**, to solve Problem (1) under the specific assumption that $g = \frac{\mu_g}{2} \|\cdot\|^2$ for some $\mu_g \geq 0$. Extending our proposed *Accelerated Proximal Alternating Predictor–Corrector* algorithm (**APAPC**) to handle a general nonsmooth function g remains an open challenge, as it would necessitate mobilizing a more complex splitting scheme such as **PD3O** or **PDDY**. Furthermore, allowing for a general g significantly complicates the landscape of convergence guarantees. For example, the linear convergence we establish in Section 3.4 for the linearly constrained setting cannot be achieved with an arbitrary function g , even if it is strongly convex [Alghunaim et al., 2021, Section VI]. Consequently, we defer the development of fully general accelerated algorithms to future work, and we focus on the structured setting with $g = \frac{\mu_g}{2} \|\cdot\|^2$, which effectively captures the key difficulties.

We note that an algorithm termed **APAPC** was previously introduced by Kovalev et al. [2020] for the specialized problem $\min f + \frac{\mu_g}{2} \|\cdot\|^2$ subject to $K^*Kx = 0$ (with $\mu_g > 0$), which arises in decentralized optimization where positive semidefinite linear constraints model network consensus. This method was later extended to general linear constraints [Salim et al., 2022a], a setting we revisit in Section 3.4. Our proposed **APAPC** substantially generalizes these previous efforts by accommodating an arbitrary nonsmooth composite term $h \circ K$. We also acknowledge the concurrent work of Zhu [2025], who proposed the *Accelerated Primal–Dual Fixed-Point* method (**APDFP**). When $\mu_g = 0$, their algorithm coincides with **APAPC**. Their analysis assumes that h is Lipschitz continuous, so that h^* has a bounded domain, to derive guarantees on the partial primal–dual gap. We do not impose such assumptions and leave the analysis of the primal–dual gap for future work.

Our main contributions are structured as follows:

- **A Unified Analysis of Nesterov Acceleration:** In Section 2, we revisit Nesterov acceleration for the minimization of $f + g$. Our refined convergence analysis of Accelerated Proximal Gradient Descent (**APGD**) offers a unified treatment of both the general convex and strongly convex regimes, achieving optimal linear convergence simply by capping the momentum parameter. Moreover, we prove weak convergence of the iterates to a solution. This framework lays the foundation for our primal–dual extension.
- **Seamless Primal-Dual Acceleration:** We propose **APAPC** for solving Problem (1) with $g = \frac{\mu_g}{2} \|\cdot\|^2$, by integrating the decoupled momentum architecture of **APGD** into the forward–

backward structure of **PAPC**. Exploiting the strong convexity of the dual problem allows for acceleration on the smooth component f while ensuring the decay of the primal–dual Lyapunov function.

- **Comprehensive Convergence Rates:** We establish accelerated convergence bounds across three distinct regimes where the dual problem is strongly convex. We prove accelerated linear convergence when the primal problem is also strongly convex ($\mu_g > 0$), and optimal $\mathcal{O}(1/t^2)$ sublinear rates otherwise. While we recover similar complexity bounds as existing algorithms in specific cases (e.g., when h is smooth), several of our results are new. Our results are summarized in the following table:

Condition on $h \circ \mathsf{K}$	h smooth	K bounded below	Linear constraint
$\mathcal{O}(1/t^2)$ convergence	Theorem 3.3	Theorem 3.8	Theorem 3.11
Linear convergence when $\mu_g > 0$	Corollary 3.4	Corollary 3.9	Corollary 3.13

- **Point Convergence:** Building upon our Lyapunov framework and recent results on Accelerated Gradient Descent, we characterize the weak convergence of the iterates for **APAPC**. To the best of our knowledge, this represents the first iterate convergence proof for an accelerated fully split primal–dual algorithm.

2 Revisiting Nesterov Acceleration

In this section, we consider the simple particular case of (1) with $h \circ \mathsf{K} = 0$:

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \Psi(x) := f(x) + g(x). \quad (5)$$

In that case, $x^* \in \mathcal{X}$ is a minimizer of Ψ if and only if it satisfies $0 \in \nabla f(x^*) + \partial g(x^*)$ [Bauschke and Combettes, 2017, Corollary 16.48]. So, we suppose that a minimizer x^* of Ψ exists and we let $\Psi^* := \Psi(x^*)$.

We consider the *Accelerated Proximal Gradient Descent* algorithm (**APGD**). Given a stepsize $\gamma > 0$ and a parameter sequence $(a_t)_{t \geq 0}$ with $a_t \geq 1$ for every $t \geq 1$, **APGD**, initialized with $z^0 = x^0$, iterates, for $t \geq 0$,

$$y^t := \left(1 - \frac{1}{a_{t+1}}\right) x^t + \frac{1}{a_{t+1}} z^t, \quad (6)$$

$$z^{t+1} := \text{prox}_{a_{t+1}\gamma g} \left(z^t - a_{t+1}\gamma \nabla f(y^t) \right), \quad (7)$$

$$x^{t+1} := \left(1 - \frac{1}{a_{t+1}}\right) x^t + \frac{1}{a_{t+1}} z^{t+1}. \quad (8)$$

APGD is a proximal extension of Nesterov’s Accelerated Gradient Descent (**AGD**) [Nesterov, 2004, p. 90], that has appeared in Auslender and Teboulle [2006, Section 5] and has been studied in Tseng [2008]. In **APGD**, γ is typically chosen as $\frac{1}{L_f}$, but $a_{t+1} \geq 1$ can be arbitrarily large. So, z^t is updated using a proximal gradient descent step with an aggressive stepsize, which is then damped by under-relaxation, as x^{t+1} is a convex combination of this new point and the previous conservative estimate x^t [Allen-Zhu and Orecchia, 2014]. Thus, x^t is a weighted average of all estimates z^0, \dots, z^t computed so far. For instance, we can check by induction that if $a_0 = 0$ and $a_t = \frac{t+1}{2}$, then $x^t = \sum_{k=1}^t \frac{2k}{t(t+1)} z^k$ for every $t \geq 1$. If x^1 is feasible, i.e., in the domain of g , then x^t

remains feasible for every $t \geq 1$, which is a desirable property. This is the case if x^0 is feasible or, as we recommend in practice, by setting $a_1 = 1$, so that $x^1 = z^1$.

We observe that if $a_{t+1} \equiv 1$, then $x^t = z^t = y^t$ and **APGD** reverts to **PGD**, which iterates $x^{t+1} := \text{prox}_{\gamma g}(x^t - \gamma \nabla f(x^t))$.

We note that **APGD** is different from the popular method **FISTA** [Beck and Teboulle, 2009]. After constructing y^t , **FISTA** updates z^t and x^t as

$$x^{t+1} := \text{prox}_{\gamma g}(y^t - \gamma \nabla f(y^t)), \quad (9)$$

$$z^{t+1} := z^t + a_{t+1}(x^{t+1} - y^t). \quad (10)$$

APGD and **FISTA** are different in the way prox_g is applied. In both methods, the variable of interest to estimate the solution is x^t , not z^t , see Remark 2.5. It may be that x^t lies on the boundary of the domain of g in **FISTA**, while it is in the interior of this domain in **APGD**. For instance, if g induces sparsity, x^t is typically more sparse in **FISTA** than in **APGD**. Explicitly promoting structure in the estimate is a property that certainly contributed to **FISTA**'s widespread adoption in inverse problems. However, after T iterations of **APGD**, one can always apply one final iteration of **PGD** to x^T . Since **PGD** is a descent method, we have $\Psi(x^{T+1}) \leq \Psi(x^T)$, and the final estimate x^{T+1} , as the output of $\text{prox}_{\gamma g}$, now has the sought properties, such as being sparse or low rank. The structural properties of the last iterate matter, and it may not be necessary to impose them at every intermediate iteration. In any case, **APGD**'s decoupling of the estimate x^t , which undergoes momentum, from the history-accumulating sequence z^t , which can accommodate dual variables, is a key property of our primal–dual extension in Section 3.

If $g = 0$ and y^t is formed as in (6), **APGD** and **FISTA** are the same and revert to **AGD**:

$$y^t := x^t + \frac{a_t - 1}{a_{t+1}}(x^t - x^{t-1}), \quad (11)$$

$$x^{t+1} := y^t - \gamma \nabla f(y^t), \quad (12)$$

with $z^t = x^t + (a_t - 1)(x^t - x^{t-1})$. We refer to Bubeck [2015] and Chambolle and Pock [2016a, Appendix B] for the convergence analysis of **AGD** and **FISTA**, see also refinements in Chambolle and Dossal [2015], Attouch and Peypouquet [2016], Aujol et al. [2025].

While the convergence properties of **APGD** have been studied [Tseng, 2008], we now provide a simple convergence analysis based on a single-step Lyapunov inequality. To the best of our knowledge, this analysis, which captures both the general convex and strongly convex regimes while decoupling μ_f and μ_g , is new.

Theorem 2.1 (Single-iteration progress of **APGD**). *Let $x^* \in \arg \min \Psi$. In **APGD**, we have*

$$\begin{aligned} \frac{1 + a_{t+1}\gamma\mu_g}{2\gamma} \|z^{t+1} - x^*\|^2 + a_{t+1}^2 (\Psi(x^{t+1}) - \Psi^*) &\leq \frac{1 - \gamma\mu_f}{2\gamma} \|z^t - x^*\|^2 + (a_{t+1}^2 - a_t) (\Psi(x^t) - \Psi^*) \\ &\quad - \left(\frac{1}{2\gamma} - \frac{L_f}{2} \right) \|z^{t+1} - z^t\|^2, \quad \forall t \geq 0. \end{aligned} \quad (13)$$

Consequently, we define the Lyapunov function

$$\mathcal{E}_{x^*}^t := \frac{1 + a_t\gamma\mu_g}{2\gamma} \|z^t - x^*\|^2 + a_t^2 (\Psi(x^t) - \Psi^*), \quad \forall t \geq 0.$$

Then, assuming that $\gamma \leq \frac{1}{L_f}$, we can ignore the last term in (13) and we have

$$\mathcal{E}_{x^*}^{t+1} \leq \max \left(\frac{1 - \gamma\mu_f}{1 + a_t\gamma\mu_g}, \frac{a_{t+1}^2 - a_{t+1}}{a_t^2} \right) \mathcal{E}_{x^*}^t, \quad \forall t \geq 0.$$

The proof of Theorem 2.1 is simple once we have the following Lemma, proved in Section 2.2. In **APGD**, we introduce, for every $t \geq 0$, the subgradient $\tilde{\nabla}g(z^{t+1}) \in \partial g(z^{t+1})$ such that $z^{t+1} + a_{t+1}\gamma\tilde{\nabla}g(z^{t+1}) = z^t - a_{t+1}\gamma\nabla f(y^t)$; it is well defined by construction of z^{t+1} in (7) [Bauschke and Combettes, 2017, Proposition 16.44].

Lemma 2.2. *In **APGD**, for every $x^* \in \arg \min \Psi$ and $t \geq 0$, we have*

$$\begin{aligned} -a_{t+1}\langle \tilde{\nabla}g(z^{t+1}) + \nabla f(y^t), z^{t+1} - x^* \rangle &\leq -a_{t+1}^2(\Psi(x^{t+1}) - \Psi^*) + (a_{t+1}^2 - a_{t+1})(\Psi(x^t) - \Psi^*) \\ &\quad - \frac{a_{t+1}\mu_g}{2}\|z^{t+1} - x^*\|^2 - \frac{\mu_f}{2}\|z^t - x^*\|^2 + \frac{L_f}{2}\|z^{t+1} - z^t\|^2 \\ &\quad - \frac{a_{t+1}}{2(L_f - \mu_f)}\|\nabla f(y^t) - \nabla f(x^*) - \mu_f(y^t - x^*)\|^2 \\ &\quad - \frac{a_{t+1}^2 - a_{t+1}}{2(L_f - \mu_f)}\|\nabla f(x^t) - \nabla f(y^t) - \mu_f(x^t - y^t)\|^2 \end{aligned}$$

(with the last two term set to zero if $\mu_f = L_f$).

Proof of Theorem 2.1. Let $x^* \in \arg \min \Psi$ and $t \geq 0$. We have $z^{t+1} - z^t = -a_{t+1}\gamma(\tilde{\nabla}g(z^{t+1}) + \nabla f(y^t))$. Therefore,

$$\begin{aligned} \|z^{t+1} - x^*\|^2 &= \|z^t - x^*\|^2 - \|z^{t+1} - z^t\|^2 + 2\langle z^{t+1} - z^t, z^{t+1} - x^* \rangle \\ &= \|z^t - x^*\|^2 - \|z^{t+1} - z^t\|^2 - 2a_{t+1}\gamma\langle \tilde{\nabla}g(z^{t+1}) + \nabla f(y^t), z^{t+1} - x^* \rangle. \end{aligned}$$

Dividing by 2γ on both sides and invoking Lemma 2.2, we obtain the inequality (13). \square

We can now focus on the choice of the sequence $(a_t)_{t \geq 0}$.

Corollary 2.3. *In the conditions of Theorem 2.1, suppose that $\gamma \leq \frac{1}{L_f}$, $a_0 = 0$, and $a_{t+1}^2 - a_{t+1} \leq a_t^2$ (equivalently, $a_{t+1} \leq \frac{1 + \sqrt{1 + 4a_t^2}}{2}$) for every $t \geq 0$ (this implies $a_1 = 1$). Then*

$$\mathcal{E}_{x^*}^t \leq \mathcal{E}_{x^*}^0 = \frac{1}{2\gamma}\|x^0 - x^*\|^2, \quad \forall t \geq 0,$$

so that

$$\Psi(x^t) - \Psi^* \leq \frac{\|x^0 - x^*\|^2}{2\gamma a_t^2}, \quad \forall t \geq 1.$$

In particular, by choosing $\gamma = \frac{1}{L_f}$ and $a_t = \frac{t+1}{2}$ (or the slightly larger sequence defined recursively by $a_{t+1} := \frac{1 + \sqrt{1 + 4a_t^2}}{2}$) for every $t \geq 1$, which satisfies the condition above, **APGD** has accelerated sublinear convergence with

$$\Psi(x^t) - \Psi^* \leq \frac{2L_f\|x^0 - x^*\|^2}{(t+1)^2}. \quad (14)$$

Moreover, suppose that $\mu_g > 0$, let $T \geq 1$, choose $\gamma = \frac{1}{L_f}$, and suppose that $a_t = a_{\sharp} := \max\left(\sqrt{\frac{L_f}{\mu_g}}, 1\right)$ for every $t \geq T$. Then $x^* = \arg \min \Psi$ is unique and **APGD** has accelerated linear convergence with

$$\mathcal{E}_{x^*}^t \leq \left(\frac{1}{1 + \sqrt{\mu_g/L_f}}\right)^{t-T} \mathcal{E}_{x^*}^T, \quad \forall t \geq T, \quad (15)$$

hence an iteration complexity (to reach an accuracy $\mathcal{E}_{x^*}^t \leq \epsilon$) $\mathcal{O}\left(\sqrt{\frac{L_f}{\mu_g}} \log \epsilon^{-1}\right)$.

These asymptotic rates are known to be optimal for first-order methods [Nesterov, 2004]. The constant 2 in (14) can be removed with the refined and optimal method **OptISTA** [Jang et al., 2025]. Fine properties of the sequence $(y_t)_{t \geq 0}$ in **APGD** have been studied in Wu et al. [2026].

Remark 2.4. In practice, we can choose $\gamma = \frac{1}{L_f}$, $a_0 = 0$ and $a_t = \min\left(\frac{t+1}{2}, a_{\sharp}\right)$ (or $a_t = \min\left(\frac{1+\sqrt{1+4a_{\sharp}^2 t}}{2}, a_{\sharp}\right)$) for every $t \geq 1$. This way, if $\mu_g > 0$, there is a first phase of about $2a_{\sharp}$ iterations where sublinear convergence of $\Psi(x^t)$ is actually faster than linear convergence, before linear convergence takes over.

Remark 2.5. In the conditions of Corollary 2.3, if $\mu_g > 0$ and $L_f \geq \mu_g$, we have linear convergence of $L_f \|z^t - x^*\|^2$ and $a_{\sharp}^2(\Psi(x^t) - \Psi^*) = \frac{L_f}{\mu_g}(\Psi(x^t) - \Psi^*)$ with same rate. Since Ψ is μ_g -strongly convex, we have $\Psi(x^t) - \Psi^* \geq \frac{\mu_g}{2} \|x^t - x^*\|^2$, so that $L_f \|x^t - x^*\|^2$ converges linearly with same rate as well. In the general convex case, $\Psi(x^t) \rightarrow \Psi^*$, but there is no guarantee that $\|z^t - x^*\|^2$ decreases. Thus, x^t is indeed the variable of interest, not z^t .

Remark 2.6. We observe in Theorem 2.1 that if $\mu_g = 0$ and $\mu_f > 0$, **APGD** converges linearly with rate at best $1 - \gamma\mu_f$, which is not accelerated. However, it is known that accelerated convergence can be obtained in this case. This is essentially achieved by transferring the strong convexity from f to g by writing $f + g = \tilde{f} + \tilde{g}$, with $\tilde{f} = f - \frac{\mu_f}{2} \|\cdot\|^2$ and $\tilde{g} = g + \frac{\mu_f}{2} \|\cdot\|^2$, and doing the corresponding changes in the algorithm.

Remark 2.7. When $\mu_g > 0$, we can get an improvement of $\sqrt{2}$ in the linear convergence rate by slightly modifying the analysis. For this, given x^* , we define $\tilde{g} = g - \frac{\mu_g}{2} \|\cdot - x^*\|^2$. Then, in the proof of Theorem 1, we have

$$\begin{aligned} \|z^{t+1} - x^*\|^2 &= \|z^t - x^*\|^2 - \|z^{t+1} - z^t\|^2 - 2a_{t+1}\gamma\mu_g \|z^{t+1} - x^*\|^2 \\ &\quad - 2a_{t+1}\gamma \langle \tilde{\nabla} \tilde{g}(z^{t+1}) + \nabla f(y^t), z^{t+1} - x^* \rangle. \end{aligned}$$

Then, invoking Lemma 2.2 with g replaced by \tilde{g} , we obtain, if $\gamma \leq \frac{1}{L_f}$,

$$\frac{1+2a_{t+1}\gamma\mu_g}{2\gamma} \|z^{t+1} - x^*\|^2 + a_{t+1}^2(\tilde{\Psi}(x^{t+1}) - \Psi^*) \leq \frac{1-\gamma\mu_f}{2\gamma} \|z^t - x^*\|^2 + (a_{t+1}^2 - a_t)(\tilde{\Psi}(x^t) - \Psi^*),$$

where $\tilde{\Psi} = f + \tilde{g} = \Psi - \frac{\mu_g}{2} \|\cdot - x^*\|^2$. Note the factor 2 in $1 + 2a_{t+1}\gamma\mu_g$, in comparison with $1 + a_{t+1}\gamma\mu_g$ in (13). Therefore, in Corollary 2.3, we can choose $a_{\sharp} := \max\left(1, \sqrt{\frac{L_f}{2\mu_g}}\right)$ and we get the contraction factor $\frac{1}{1+\sqrt{2\mu_g/L_f}}$ on the modified Lyapunov function. Asymptotically, this is the best known rate for strongly convex composite problems [Ushiyama, 2025].

2.1 Point Convergence of APGD

We complement the analysis of APGD by establishing its point convergence. Historically, proving convergence of the iterates for accelerated methods has been a notoriously difficult open problem. In a decisive paper, Chambolle and Dossal [2015] proved that if one chooses the momentum parameter in FISTA as $a_t = \frac{t+\beta-1}{\beta}$ for some $\beta > 2$, the sequence $(x^t)_{t \geq 0}$ converges weakly to a solution $x^* \in \arg \min \Psi$. However, the convergence of FISTA, and even its special case AGD, in the critical case $a_t = \frac{t+1}{2}$ remained unresolved for another decade. Recently, in late 2025, a breakthrough was achieved simultaneously by two independent teams: Jang and Ryu [2025] proved the convergence of AGD (assuming \mathcal{X} is finite-dimensional), while Boç et al. [2025b] established the weak convergence of FISTA, under the general conditions $a_{t+1}^2 - a_{t+1} \leq a_t^2$ and $a_t \rightarrow \infty$. With minor adjustments to the proof technique from Jang and Ryu [2025] to accommodate infinite-dimensional Hilbert spaces, we demonstrate that this fundamental result also applies to APGD.

Theorem 2.8. *In APGD, suppose that $\gamma \leq \frac{1}{L_f}$, $a_0 = 0$, $a_{t+1}^2 - a_{t+1} \leq a_t^2$ for every $t \geq 0$, and $a_t \rightarrow \infty$. Then $(z^t)_{t \geq 0}$ is bounded, and there exists $x^* \in \arg \min \Psi$ such that $(x^t)_{t \geq 0}$ and $(y^t)_{t \geq 0}$ both converge weakly to x^* .*

Proof. For every $x^* \in \arg \min \Psi$ and $t \geq 0$, we define the Lyapunov function $\mathcal{E}_{x^*}^t$ as in Theorem 2.1, with $\mu_f = \mu_g = 0$; then from Corollary 2.3, we have $\frac{1}{2\gamma} \|z^t - x^*\|^2 \leq \mathcal{E}_{x^*}^t \leq \mathcal{E}_{x^*}^0$. Therefore, the sequence $(z^t)_{t \geq 0}$ is bounded and there exists $M > 0$ such that $\|z^t\| \leq M$ for every $t \geq 0$. Consequently, it follows from (8) that $(x^t)_{t \geq 0}$ is also bounded, because $x^0 = z^0$ and $\|x^{t+1}\| \leq \left(1 - \frac{1}{a_{t+1}}\right) \|x^t\| + \frac{1}{a_{t+1}} \|z^{t+1}\| \leq \max(\|x^t\|, M) \leq \dots \leq \max(\|x^0\|, M) = M$. Therefore, $(x^t)_{t \geq 0}$ has at least one weak cluster point. Moreover, since $a_t \rightarrow \infty$ and $a_t^2(\Psi(x^t) - \Psi^*) \leq \mathcal{E}^0$, we have $\Psi(x^t) \rightarrow \Psi^*$. Since Ψ is convex and lower semicontinuous, it is weakly sequentially lower semicontinuous; consequently, every weak cluster point x^∞ of $(x^t)_{t \geq 0}$ satisfies $\Psi(x^\infty) \leq \liminf_{t \rightarrow \infty} \Psi(x^t) = \Psi^*$, and hence is a minimizer of Ψ .

Let $\hat{x}^* \in \arg \min \Psi$ and $\check{x}^* \in \arg \min \Psi$ be two weak cluster points of $(x^t)_{t \geq 0}$. We want to show that $\hat{x}^* = \check{x}^*$. To that aim, we define, for every $t \geq 0$, $\mathcal{A}^t := \|x^t - \hat{x}^*\|^2 - \|x^t - \check{x}^*\|^2$ and $\mathcal{B}^t := 2\gamma(\mathcal{E}_{\hat{x}^*}^t - \mathcal{E}_{\check{x}^*}^t) = \|z^t - \hat{x}^*\|^2 - \|z^t - \check{x}^*\|^2$. Since the sequences $(\mathcal{E}_{\hat{x}^*}^t)_{t \geq 0}$ and $(\mathcal{E}_{\check{x}^*}^t)_{t \geq 0}$ are nonnegative and nonincreasing, they both converge, so that $(\mathcal{B}^t)_{t \geq 0}$ converges to some limit $\mathcal{B}^\infty \in \mathbb{R}$. Moreover, for every $t \geq 0$, we can expand the norms to obtain

$$\begin{aligned} \mathcal{B}^{t+1} &= -2\langle z^{t+1}, \hat{x}^* - \check{x}^* \rangle + \|\hat{x}^*\|^2 - \|\check{x}^*\|^2, \\ \mathcal{A}^{t+1} &= -2\langle x^{t+1}, \hat{x}^* - \check{x}^* \rangle + \|\hat{x}^*\|^2 - \|\check{x}^*\|^2, \\ \mathcal{A}^t &= -2\langle x^t, \hat{x}^* - \check{x}^* \rangle + \|\hat{x}^*\|^2 - \|\check{x}^*\|^2. \end{aligned}$$

Using (8) again, we get $z^{t+1} = x^{t+1} + (a_{t+1} - 1)(x^{t+1} - x^t)$, so that $\mathcal{B}^{t+1} = \mathcal{A}^{t+1} + (a_{t+1} - 1)(\mathcal{A}^{t+1} - \mathcal{A}^t)$. Furthermore, for every $t \geq 0$ the condition $a_{t+1}^2 - a_{t+1} \leq a_t^2$ implies $a_{t+1} \leq \frac{1 + \sqrt{1 + 4a_t^2}}{2} \leq \frac{1 + \sqrt{1} + \sqrt{4a_t^2}}{2} = a_t + 1$ and, by recursion, $a_t \leq t$. Since $a_t \rightarrow \infty$, there exists $T \geq 2$ such that $a_t > 1$ for every $t \geq T$, and we have $\sum_{t=T}^{\infty} \frac{1}{a_{t-1}} \geq \sum_{t=T}^{\infty} \frac{1}{t-1} = \infty$. The conditions are therefore met to invoke Lemma A.4 in Boç et al. [2025a] (see another proof of this result in Bauschke and Moursi [2026]), and we conclude that $\mathcal{A}^t \rightarrow \mathcal{B}^\infty$. Finally, let $(x^{t_j})_{j \geq 0}$ and $(x^{t'_j})_{j \geq 0}$ be two subsequences of $(x^t)_{t \geq 0}$ converging weakly to \hat{x}^* and \check{x}^* , respectively. Then $\mathcal{A}^{t_j} \rightarrow -\|\hat{x}^* - \check{x}^*\|^2$ and $\mathcal{A}^{t'_j} \rightarrow \|\check{x}^* - \hat{x}^*\|^2$. Because the entire sequence $(\mathcal{A}^t)_{t \geq 0}$ converges to a single limit, these subsequential limits must

coincide. Hence, $\mathcal{B}^\infty = -\|\hat{x}^* - \check{x}^*\|^2 = \|\hat{x}^* - \check{x}^*\|^2$, which implies $\hat{x}^* = \check{x}^*$. It follows that $(x^t)_{t \geq 0}$ has exactly one weak cluster point x^* , and hence converges weakly to x^* .

Using (6), we have $\|y^t - x^t\| = \frac{1}{a_{t+1}}\|x^t - z^t\| \rightarrow 0$, because $a_t \rightarrow \infty$ and $(x^t - z^t)_{t \geq 0}$ is bounded. Hence, $(y^t)_{t \geq 0}$ also converges weakly to x^* . \square

2.2 Proof of Lemma 2.2

Let $x^* \in \arg \min \Psi$ and $t \geq 0$. We define the Bregman distances

$$\begin{aligned} D_f : x \in \mathcal{X} &\mapsto f(x) - f(x^*) - \langle x - x^*, \nabla f(x^*) \rangle \geq 0, \\ D_g : x \in \mathcal{X} &\mapsto g(x) - g(x^*) - \langle x - x^*, \tilde{\nabla} g(x^*) \rangle \geq 0, \end{aligned}$$

where $\tilde{\nabla} g(x^*) := -\nabla f(x^*)$. We want to express

$$\begin{aligned} \langle \tilde{\nabla} g(z^{t+1}) + \nabla f(y^t), z^{t+1} - x^* \rangle &= \langle \tilde{\nabla} g(z^{t+1}) - \tilde{\nabla} g(x^*), z^{t+1} - x^* \rangle \\ &\quad + \langle \nabla f(y^t) - \nabla f(x^*), z^{t+1} - x^* \rangle \end{aligned}$$

using D_f and D_g .

Step 1: Bounding the nonsmooth term. By μ_g -strong convexity of g , we have $g(x^*) \geq g(z^{t+1}) + \langle \tilde{\nabla} g(z^{t+1}), x^* - z^{t+1} \rangle + \frac{\mu_g}{2} \|z^{t+1} - x^*\|^2$, so that $D_g(z^{t+1}) \leq \langle \tilde{\nabla} g(z^{t+1}) - \tilde{\nabla} g(x^*), z^{t+1} - x^* \rangle - \frac{\mu_g}{2} \|z^{t+1} - x^*\|^2$, and

$$\begin{aligned} &-a_{t+1} \langle \tilde{\nabla} g(z^{t+1}) - \tilde{\nabla} g(x^*), z^{t+1} - x^* \rangle \\ &\leq -a_{t+1} D_g(z^{t+1}) - \frac{a_{t+1} \mu_g}{2} \|z^{t+1} - x^*\|^2 \\ &\leq -a_{t+1}^2 D_g(x^{t+1}) + (a_{t+1}^2 - a_{t+1}) D_g(x^t) - \frac{a_{t+1} \mu_g}{2} \|z^{t+1} - x^*\|^2. \end{aligned} \tag{16}$$

The last inequality follows from the convexity of D_g : we have $D_g(x^{t+1}) \leq (1 - \frac{1}{a_{t+1}}) D_g(x^t) + \frac{1}{a_{t+1}} D_g(z^{t+1})$, so that $D_g(z^{t+1}) \geq a_{t+1} D_g(x^{t+1}) - (a_{t+1} - 1) D_g(x^t)$.

Step 2: Expanding the gradient term. We expand $\langle \nabla f(y^t) - \nabla f(x^*), z^{t+1} - x^* \rangle$ into inner products of the form $\langle \nabla f(y^t) - \nabla f(x^*), y^t - X \rangle$, where X is x^* , x^t , or x^{t+1} , and subsequently use convexity and smoothness at y^t to express them using $D_f(y^t)$, $D_f(x^t)$, and $D_f(x^{t+1})$. From $y^t = \frac{1}{a_{t+1}} z^t + (1 - \frac{1}{a_{t+1}}) x^t$, we obtain $z^t - x^* = (y^t - x^*) + (a_{t+1} - 1)(y^t - x^t)$. Moreover, $x^{t+1} - y^t = \frac{1}{a_{t+1}}(z^{t+1} - z^t)$. Hence,

$$\begin{aligned} \langle \nabla f(y^t) - \nabla f(x^*), z^{t+1} - x^* \rangle &= \langle \nabla f(y^t) - \nabla f(x^*), z^{t+1} - z^t \rangle \\ &\quad + \langle \nabla f(y^t) - \nabla f(x^*), z^t - x^* \rangle \\ &= -a_{t+1} \langle \nabla f(y^t) - \nabla f(x^*), y^t - x^{t+1} \rangle \\ &\quad + \langle \nabla f(y^t) - \nabla f(x^*), y^t - x^* \rangle \\ &\quad + (a_{t+1} - 1) \langle \nabla f(y^t) - \nabla f(x^*), y^t - x^t \rangle. \end{aligned} \tag{17}$$

Step 2-a: Using smoothness at y^t and x^{t+1} . Since f is L_f -smooth, we have

$$\begin{aligned} f(x^{t+1}) - f(x^*) &\leq f(y^t) - f(x^*) + \langle \nabla f(y^t), x^{t+1} - y^t \rangle + \frac{L_f}{2} \|x^{t+1} - y^t\|^2 \\ &= f(y^t) - f(x^*) + \langle \nabla f(y^t) - \nabla f(x^*), x^{t+1} - y^t \rangle \\ &\quad + \langle \nabla f(x^*), x^{t+1} - x^* \rangle - \langle \nabla f(x^*), y^t - x^* \rangle + \frac{L_f}{2} \|x^{t+1} - y^t\|^2, \end{aligned}$$

so that $D_f(x^{t+1}) - D_f(y^t) - \frac{L_f}{2} \|x^{t+1} - y^t\|^2 \leq \langle \nabla f(y^t) - \nabla f(x^*), x^{t+1} - y^t \rangle$.

Step 2-b: Using strong convexity and smoothness at y^t and x^t . Since f is μ_f -strongly convex and L_f -smooth, we have [Taylor et al., 2017, Theorem 4]

$$\begin{aligned} f(x^t) - f(x^*) &\geq f(y^t) - f(x^*) + \langle \nabla f(y^t), x^t - y^t \rangle + \frac{\mu_f}{2} \|x^t - y^t\|^2 \\ &\quad + \frac{1}{2(L_f - \mu_f)} \|\nabla f(x^t) - \nabla f(y^t) - \mu_f(x^t - y^t)\|^2 \\ &= f(y^t) - f(x^*) + \langle \nabla f(y^t) - \nabla f(x^*), x^t - y^t \rangle \\ &\quad + \langle \nabla f(x^*), x^t - x^* \rangle - \langle \nabla f(x^*), y^t - x^* \rangle + \frac{\mu_f}{2} \|y^t - x^t\|^2 \\ &\quad + \frac{1}{2(L_f - \mu_f)} \|\nabla f(x^t) - \nabla f(y^t) - \mu_f(x^t - y^t)\|^2 \end{aligned}$$

(with the last term set to zero if $\mu_f = L_f$), so that $\langle \nabla f(y^t) - \nabla f(x^*), y^t - x^t \rangle \geq -D_f(x^t) + D_f(y^t) + \frac{\mu_f}{2} \|y^t - x^t\|^2 + \frac{1}{2(L_f - \mu_f)} \|\nabla f(x^t) - \nabla f(y^t) - \mu_f(x^t - y^t)\|^2$.

Step 2-c: Using strong convexity and smoothness at y^t and x^* . Using the same derivations as in Step 2-b, with x^t replaced by x^* , we obtain $\langle \nabla f(y^t) - \nabla f(x^*), y^t - x^* \rangle \geq D_f(y^t) + \frac{\mu_f}{2} \|y^t - x^*\|^2 + \frac{1}{2(L_f - \mu_f)} \|\nabla f(y^t) - \nabla f(x^*) - \mu_f(y^t - x^*)\|^2$.

Therefore, using the assumption $a_{t+1} \geq 1$, so that $a_{t+1} - 1 \geq 0$ in (17), we have

$$\begin{aligned} -a_{t+1} \langle \nabla f(y^t) - \nabla f(x^*), z^{t+1} - x^* \rangle &\leq a_{t+1}^2 \left(-D_f(x^{t+1}) + D_f(y^t) + \frac{L_f}{2} \|x^{t+1} - y^t\|^2 \right) \\ &\quad - a_{t+1} \left(D_f(y^t) + \frac{\mu_f}{2} \|y^t - x^*\|^2 \right) \\ &\quad + \frac{1}{2(L_f - \mu_f)} \|\nabla f(y^t) - \nabla f(x^*) - \mu_f(y^t - x^*)\|^2 \\ &\quad - (a_{t+1}^2 - a_{t+1}) \left(-D_f(x^t) + D_f(y^t) + \frac{\mu_f}{2} \|y^t - x^t\|^2 \right) \\ &\quad + \frac{1}{2(L_f - \mu_f)} \|\nabla f(x^t) - \nabla f(y^t) - \mu_f(x^t - y^t)\|^2 \\ &= -a_{t+1}^2 D_f(x^{t+1}) + (a_{t+1}^2 - a_{t+1}) D_f(x^t) + \frac{L_f}{2} \|z^{t+1} - z^t\|^2 \\ &\quad - a_{t+1} \frac{\mu_f}{2} \|y^t - x^*\|^2 - (a_{t+1}^2 - a_{t+1}) \frac{\mu_f}{2} \|y^t - x^t\|^2 \\ &\quad - \frac{a_{t+1}}{2(L_f - \mu_f)} \|\nabla f(y^t) - \nabla f(x^*) - \mu_f(y^t - x^*)\|^2 \\ &\quad - \frac{a_{t+1}^2 - a_{t+1}}{2(L_f - \mu_f)} \|\nabla f(x^t) - \nabla f(y^t) - \mu_f(x^t - y^t)\|^2. \end{aligned}$$

Furthermore, by Jensen's inequality, from the relation $\frac{1}{a_{t+1}}(z^t - x^*) = \frac{1}{a_{t+1}}(y^t - x^*) + \frac{a_{t+1}-1}{a_{t+1}}(y^t - x^t)$, we have $\left\| \frac{1}{a_{t+1}}(z^t - x^*) \right\|^2 \leq \frac{1}{a_{t+1}} \|y^t - x^*\|^2 + \frac{a_{t+1}-1}{a_{t+1}} \|y^t - x^t\|^2$, so that $\|z^t - x^*\|^2 \leq a_{t+1} \|y^t - x^*\|^2 +$

Algorithm 1 Accelerated Proximal Alternating Predictor–Corrector algorithm (APAPC)

- 1: **input:** primal stepsize $\gamma > 0$, dual stepsize $\tau > 0$, parameter sequence $(a_t)_{t \geq 0}$ with $a_t \geq 1$
 $\forall t \geq 1$, initial primal estimate $x^0 = z^0 \in \mathcal{X}$, initial dual estimate $u^0 = v^0 \in \mathcal{U}$
 - 2: **for** $t = 0, 1, \dots$ **do**
 - 3: $y^t := \left(1 - \frac{1}{a_{t+1}}\right) x^t + \frac{1}{a_{t+1}} z^t$
 - 4: $\hat{z}^t := (1 + a_{t+1} \gamma \mu_g)^{-1} (z^t - a_{t+1} \gamma \nabla f(y^t) - a_{t+1} \gamma \mathbf{K}^* v^t)$
 - 5: $v^{t+1} := \text{prox}_{\frac{\tau}{a_{t+1}} h^*} \left(v^t + \frac{\tau}{a_{t+1}} \mathbf{K} \hat{z}^t \right)$
 - 6: $z^{t+1} := (1 + a_{t+1} \gamma \mu_g)^{-1} (z^t - a_{t+1} \gamma \nabla f(y^t) - a_{t+1} \gamma \mathbf{K}^* v^{t+1})$
 - 7: $x^{t+1} := \left(1 - \frac{1}{a_{t+1}}\right) x^t + \frac{1}{a_{t+1}} z^{t+1}$
 - 8: $u^{t+1} := \left(1 - \frac{1}{a_{t+1}}\right) u^t + \frac{1}{a_{t+1}} v^{t+1}$ (optional, used only in the analysis)
 - 9: **end for**
-

$(a_{t+1}^2 - a_{t+1}) \|y^t - x^t\|^2$. Hence,

$$\begin{aligned}
-a_{t+1} \left\langle \nabla f(y^t) - \nabla f(x^*), z^{t+1} - x^* \right\rangle &\leq -a_{t+1}^2 D_f(x^{t+1}) + (a_{t+1}^2 - a_{t+1}) D_f(x^t) \\
&\quad + \frac{L_f}{2} \|z^{t+1} - z^t\|^2 - \frac{\mu_f}{2} \|z^t - x^*\|^2 \\
&\quad - \frac{a_{t+1}^2 - a_{t+1}}{2(L_f - \mu_f)} \|\nabla f(x^t) - \nabla f(y^t) - \mu_f(x^t - y^t)\|^2. \tag{18}
\end{aligned}$$

Combining (16) and (18), and noting that $D_f + D_g = f - f(x^*) + g - g(x^*) = \Psi - \Psi^*$, we obtain the stated inequality.

3 The Accelerated Proximal Alternating Predictor–Corrector Algorithm (APAPC)

We now consider Problem (1) under the specific assumption that $g := \frac{\mu_g}{2} \|\cdot\|^2$:

$$\underset{x \in \mathcal{X}}{\text{minimize}} \Psi(x) := f(x) + \frac{\mu_g}{2} \|x\|^2 + h(\mathbf{K}x). \tag{19}$$

For simplicity, we assume that $L_f \geq \mu_g$; otherwise, acceleration is unnecessary. This problem is more general than Problem (5): setting $\mu_g = 0$ and $\mathbf{K} = \mathbf{I}$ recovers the minimization of $f + h$, as detailed in Remark 3.10. The corresponding dual problem is

$$\underset{u \in \mathcal{U}}{\text{minimize}} \left(f + \frac{\mu_g}{2} \|\cdot\|^2 \right)^* (-\mathbf{K}^* u) + h^*(u). \tag{20}$$

To solve this, we introduce the *Accelerated Proximal Alternating Predictor–Corrector* algorithm (APAPC), given in Algorithm 1. It proceeds as follows. The three core steps (4)–(6) mirror the PAPC mechanism for updating z^t and v^t , with specific stepsizes and the key distinction that ∇f is evaluated at y^t rather than at z^t . The first step (4) predicts the updated primal variable \hat{z}^t using (z^t, v^t) , the second step (5) performs the dual update using (\hat{z}^t, v^t) , and the third step (6) corrects the primal variable using the updated dual variable v^{t+1} . The final steps (7)–(8) form the primal and dual estimates x^t and u^t as weighted averages of the past iterates z^k and v^k , $k = 0, \dots, t$.

The intermediate point y^t is also a weighted average of the past z^k , mirroring the momentum architecture of **APGD**.

When $a_t \equiv 1$ and $\mu_g = 0$, we have $x^t = y^t = z^t$ and $v^t = u^t$, so that **APAPC** reduces exactly to **PAPC** (see (4)), which justifies its name. Adding the function $g := \frac{\mu_g}{2} \|\cdot\|^2$, whose proximity operator is a simple scaling, to Problem (19) does not alter the forward–backward structure of **PAPC** [Condat et al., 2023, Section 4.1]. However, this term, when positive, is crucial for achieving accelerated linear convergence, as demonstrated for **APGD** in Section 2.

Thus, **APAPC** addresses the challenge of accelerating **PAPC** by organically combining the decoupled momentum architecture of **APGD** (exhibited in Section 2) with the forward–backward splitting structure of **PAPC**. The result is a fully split, single-loop algorithm that accommodates the composite term $h \circ \mathsf{K}$ while fully preserving optimal acceleration on the smooth component f .

In the remainder of this section, we establish accelerated $\mathcal{O}(1/t^2)$ convergence of **APAPC** in three settings where the dual problem is strongly convex. If, in addition, $\mu_g > 0$, then the primal problem is also strongly convex and **APAPC** achieves accelerated linear convergence. We mention that the symmetric case where only the primal problem is strongly convex (i.e., $\mu_g > 0$ and $h \circ \mathsf{K}$ is arbitrary) can be handled by disabling momentum and using a *decreasing* parameter a_t in Steps (4)–(6) [Condat et al., 2022, 2025]. We do not consider this case here, as our focus is on Nesterov acceleration of the smooth term, which relies on strong convexity of the dual problem.

We assume $a_t \geq 1$ for every $t \geq 1$ and $a_0 \geq 0$. We set $a_1 = 1$ in our convergence results, ensuring that the first iteration of **APAPC** coincides with a **PAPC** iteration. In particular, $x^1 = z^1$ and $u^1 = v^1$ belongs to the domain of h^* , as do all subsequent iterates u^t .

3.1 General Lyapunov Analysis

Following the approach used for **APGD** in Theorem 2.1, we first establish a Lyapunov inequality for **APAPC**. In the purely primal setting of Section 2, convergence rates were characterized by the primal gap $\Psi(x^t) - \Psi(x^*)$. In the primal–dual setting, we instead consider the Lagrangian

$$\mathcal{L} : (x, u) \in \mathcal{X} \times \mathcal{U} \mapsto (f + g)(x) + \langle \mathsf{K}x, u \rangle - h^*(u).$$

For every saddle-point solution (x^*, u^*) to (3), we define the Lagrangian gap as

$$\mathcal{G}_{x^*, u^*} := \mathcal{L}(x, u^*) - \mathcal{L}(x^*, u) \geq \frac{\mu_g}{2} \|x - x^*\|^2 + \frac{\mu_{h^*}}{2} \|u - u^*\|^2, \quad \forall (x, u) \in \mathcal{X} \times \mathcal{U}.$$

When $h = 0$, the dual variable can be fixed to zero, and this gap reduces to the primal objective gap: $\mathcal{L}(x, 0) - \mathcal{L}(x^*, 0) = (f + g)(x) - (f + g)(x^*) = \Psi(x) - \Psi(x^*)$. Thus, the Lagrangian gap provides a natural extension of the primal objective gap to the primal–dual setting. However, the Lagrangian gap alone may not serve as a fully informative optimality measure in all problem instances, as the condition $\mathcal{G}_{x^*, u^*}(x, u) = 0$ does not generally guarantee that (x, u) is a primal–dual solution pair. This limitation underscores the importance of complementing the convergence rate analysis with a study of the iterate convergence, which we undertake in Section 3.5.

To state our core single-iteration inequality for **APAPC**, we define, for every $t \geq 0$, the subgradient $\tilde{\nabla} h^*(v^{t+1}) \in \partial h^*(v^{t+1})$ satisfying $v^{t+1} + \frac{\tau}{a_{t+1}} \tilde{\nabla} h^*(v^{t+1}) = v^t + \frac{\tau}{a_{t+1}} \mathsf{K} \hat{z}^t$.

Theorem 3.1 (Single-iteration progress of APAPC). *Let (x^*, u^*) be a solution of (3). Then, for every $t \geq 0$, the iterates of APAPC satisfy*

$$\begin{aligned}
& \frac{1 + a_{t+1}\gamma\mu_g}{2\gamma} \|z^{t+1} - x^*\|^2 + \frac{a_{t+1}^2 + a_{t+1}\tau\mu_h^*}{2\tau} \|v^{t+1} - u^*\|^2 - \frac{a_{t+1}^2\gamma}{2 + 2a_{t+1}\gamma\mu_g} \|\mathbf{K}^*(v^{t+1} - u^*)\|^2 \\
& \quad + a_{t+1}^2 \mathcal{G}_{x^*, u^*}(x^{t+1}, u^{t+1}) \\
& \leq \frac{1}{2\gamma} \|z^t - x^*\|^2 + \frac{a_{t+1}^2}{2\tau} \|v^t - u^*\|^2 - \frac{a_{t+1}^2\gamma}{2 + 2a_{t+1}\gamma\mu_g} \|\mathbf{K}^*(v^t - u^*)\|^2 \\
& \quad - \left(\frac{1}{2\gamma} - \frac{L_f}{2} \right) \|z^{t+1} - z^t\|^2 - \frac{\tau}{2} \|\mathbf{K}z^{t+1} - \tilde{\nabla}h^*(v^{t+1})\|^2 \\
& \quad + (a_{t+1}^2 - a_{t+1}) \mathcal{G}_{x^*, u^*}(x^t, u^t) - \frac{a_{t+1}}{2L_f} \|\nabla f(y^t) - \nabla f(x^*)\|^2 - \frac{a_{t+1}^2 - a_{t+1}}{2L_f} \|\nabla f(x^t) - \nabla f(y^t)\|^2.
\end{aligned} \tag{21}$$

Proof. Let (x^*, u^*) be a solution of (3). We conduct our analysis in the product space $\mathcal{X} \times \mathcal{U}$, adopting matrix-vector notation for the variables and linear operators. For every $t \geq 0$, we define the block linear operator

$$\mathbf{Q}^t := \begin{bmatrix} \frac{1}{\gamma} & 0 \\ 0 & \mathbf{Q}_{\mathcal{U}}^t \end{bmatrix}, \quad \text{with } \mathbf{Q}_{\mathcal{U}}^t := \frac{a_t^2}{\tau} \mathbf{I} - \frac{a_t^2\gamma}{1 + a_t\gamma\mu_g} \mathbf{K}\mathbf{K}^*. \tag{22}$$

We define the pseudo inner product $\langle \cdot, \cdot \rangle_{\mathbf{Q}^t} = \langle \cdot, \mathbf{Q}^t \cdot \rangle$ on $\mathcal{X} \times \mathcal{U}$. This choice is motivated by the fact that PABC can be cast as a preconditioned forward-backward algorithm in $\mathcal{X} \times \mathcal{U}$ endowed with this metric [Combettes et al., 2014, Condat et al., 2023].

We define $\tilde{\nabla}h^*(u^*) := \mathbf{K}x^*$. Let $t \geq 0$. From Steps (4) and (6), we have $(1 + a_{t+1}\gamma\mu_g)(z^{t+1} - \hat{z}^t) = -a_{t+1}\gamma\mathbf{K}^*(v^{t+1} - v^t)$. Substituting this relation into the dual update yields $\mathbf{Q}_{\mathcal{U}}^{t+1}(v^{t+1} - v^t) = \frac{a_{t+1}^2}{\tau}(v^{t+1} - v^t) - \frac{a_{t+1}^2\gamma}{1 + a_{t+1}\gamma\mu_g} \mathbf{K}\mathbf{K}^*(v^{t+1} - v^t) = a_{t+1}\mathbf{K}\hat{z}^t - a_{t+1}\tilde{\nabla}h^*(v^{t+1}) + a_{t+1}\mathbf{K}(z^{t+1} - \hat{z}^t) = a_{t+1}\mathbf{K}z^{t+1} - a_{t+1}\tilde{\nabla}h^*(v^{t+1})$. Therefore,

$$\begin{aligned}
\mathbf{Q}^{t+1} \begin{bmatrix} z^{t+1} - z^t \\ v^{t+1} - v^t \end{bmatrix} &= a_{t+1} \begin{bmatrix} -\mu_g z^{t+1} - \nabla f(y^t) - \mathbf{K}^*v^{t+1} \\ \mathbf{K}z^{t+1} - \tilde{\nabla}h^*(v^{t+1}) \end{bmatrix} \\
&= a_{t+1} \begin{bmatrix} -\mu_g(z^{t+1} - x^*) - \nabla f(y^t) + \nabla f(x^*) - \mathbf{K}^*(v^{t+1} - u^*) \\ \mathbf{K}(z^{t+1} - x^*) - \tilde{\nabla}h^*(v^{t+1}) + \tilde{\nabla}h^*(u^*) \end{bmatrix},
\end{aligned}$$

where we used the optimality condition $0 = \mu_g x^* + \nabla f(x^*) + \mathbf{K}^*u^*$. Hence, expanding the norm

gives

$$\begin{aligned}
\frac{1}{2} \left\| \begin{bmatrix} z^{t+1} - x^* \\ v^{t+1} - u^* \end{bmatrix} \right\|_{\mathbf{Q}^{t+1}}^2 &= \frac{1}{2} \left\| \begin{bmatrix} z^t - x^* \\ v^t - u^* \end{bmatrix} \right\|_{\mathbf{Q}^{t+1}}^2 - \frac{1}{2} \left\| \begin{bmatrix} z^{t+1} - z^t \\ v^{t+1} - v^t \end{bmatrix} \right\|_{\mathbf{Q}^{t+1}}^2 \\
&\quad + \left\langle \begin{bmatrix} z^{t+1} - z^t \\ v^{t+1} - v^t \end{bmatrix}, \begin{bmatrix} z^{t+1} - x^* \\ v^{t+1} - u^* \end{bmatrix} \right\rangle_{\mathbf{Q}^{t+1}} \\
&= \frac{1}{2} \left\| \begin{bmatrix} z^t - x^* \\ v^t - u^* \end{bmatrix} \right\|_{\mathbf{Q}^{t+1}}^2 - \frac{1}{2\gamma} \|z^{t+1} - z^t\|^2 \\
&\quad - \frac{1}{2} \langle v^{t+1} - v^t, \mathbf{Q}_{\mathcal{U}}^t(v^{t+1} - v^t) \rangle \\
&\quad - a_{t+1}\mu_g \|z^{t+1} - x^*\|^2 - a_{t+1} \langle \nabla f(y^t) - \nabla f(x^*), z^{t+1} - x^* \rangle \\
&\quad - a_{t+1} \langle \tilde{\nabla} h^*(v^{t+1}) - \tilde{\nabla} h^*(u^*), v^{t+1} - u^* \rangle \\
&\quad + a_{t+1} \underbrace{\left\langle \begin{bmatrix} -\mathbf{K}^*(v^{t+1} - u^*) \\ \mathbf{K}(z^{t+1} - x^*) \end{bmatrix}, \begin{bmatrix} z^{t+1} - x^* \\ v^{t+1} - u^* \end{bmatrix} \right\rangle}_{0}.
\end{aligned} \tag{23}$$

Moreover,

$$\begin{aligned}
\langle v^{t+1} - v^t, \mathbf{Q}_{\mathcal{U}}^t(v^{t+1} - v^t) \rangle &\geq \frac{1}{\|\mathbf{Q}_{\mathcal{U}}^t\|} \|\mathbf{Q}_{\mathcal{U}}^t(v^{t+1} - v^t)\|^2 \\
&\geq \frac{\tau}{a_{t+1}^2} \|\mathbf{Q}_{\mathcal{U}}^t(v^{t+1} - v^t)\|^2 \\
&= \tau \|\mathbf{K}z^{t+1} - \tilde{\nabla} h^*(v^{t+1})\|^2.
\end{aligned}$$

We note that if $\|\mathbf{Q}_{\mathcal{U}}^t\| = 0$, this inequality holds trivially, as both sides vanish and $\mathbf{K}z^{t+1} - \tilde{\nabla} h^*(v^{t+1}) = 0$.

We introduce the Bregman distances

$$\begin{aligned}
D_f &: x \in \mathcal{X} \mapsto f(x) - f(x^*) - \langle x - x^*, \nabla f(x^*) \rangle \geq 0, \\
D_g &: x \in \mathcal{X} \mapsto g(x) - g(x^*) - \langle x - x^*, \mu_g x^* \rangle = \frac{\mu_g}{2} \|x - x^*\|^2 \geq 0, \\
D_{h^*} &: u \in \mathcal{U} \mapsto h^*(u) - h^*(u^*) - \langle u - u^*, \tilde{\nabla} h^*(u^*) \rangle \geq 0.
\end{aligned}$$

Then, the Lagrangian gap can be expressed as

$$\mathcal{G}_{x^*, u^*}(x, u) = \mathcal{L}(x, u^*) - \mathcal{L}(x^*, u) = (D_f + D_g)(x) + D_{h^*}(u), \quad \forall (x, u) \in \mathcal{X} \times \mathcal{U}. \tag{24}$$

Proceeding as in the proof of Lemma 2.2, we obtain

$$\begin{aligned}
&- a_{t+1}\mu_g \|z^{t+1} - x^*\|^2 - a_{t+1} \langle \nabla f(y^t) - \nabla f(x^*), z^{t+1} - x^* \rangle \\
&\leq -a_{t+1}^2(D_f + D_g)(x^{t+1}) + (a_{t+1}^2 - a_{t+1})(D_f + D_g)(x^t) - \frac{a_{t+1}\mu_g}{2} \|z^{t+1} - x^*\|^2 + \frac{L_f}{2} \|z^{t+1} - z^t\|^2 \\
&\quad - \frac{a_{t+1}}{2L_f} \|\nabla f(y^t) - \nabla f(x^*)\|^2 - \frac{a_{t+1}^2 - a_{t+1}}{2L_f} \|\nabla f(x^t) - \nabla f(y^t)\|^2.
\end{aligned}$$

Similarly, by the μ_{h^*} -strong convexity of h^* , we have $h^*(u^*) \geq h^*(v^{t+1}) + \langle \tilde{\nabla} h^*(v^{t+1}), u^* - v^{t+1} \rangle + \frac{\mu_{h^*}}{2} \|v^{t+1} - u^*\|^2$, so that $D_{h^*}(v^{t+1}) \leq \langle \tilde{\nabla} h^*(v^{t+1}) - \tilde{\nabla} h^*(u^*), v^{t+1} - u^* \rangle - \frac{\mu_{h^*}}{2} \|v^{t+1} - u^*\|^2$. Moreover, by the convexity of D_{h^*} , we have $D_{h^*}(u^{t+1}) \leq (1 - \frac{1}{a_{t+1}})D_{h^*}(u^t) + \frac{1}{a_{t+1}}D_{h^*}(v^{t+1})$. Hence,

$$\begin{aligned} & -a_{t+1} \langle \tilde{\nabla} h^*(v^{t+1}) - \tilde{\nabla} h^*(u^*), v^{t+1} - u^* \rangle \\ & \leq -a_{t+1} D_{h^*}(v^{t+1}) - \frac{a_{t+1} \mu_{h^*}}{2} \|v^{t+1} - u^*\|^2 \\ & \leq -a_{t+1}^2 D_{h^*}(u^{t+1}) + (a_{t+1}^2 - a_{t+1}) D_{h^*}(u^t) - \frac{a_{t+1} \mu_{h^*}}{2} \|v^{t+1} - u^*\|^2. \end{aligned}$$

Substituting these bounds into (23) and expanding the squared norms, we obtain

$$\begin{aligned} & \frac{1}{2\gamma} \|z^{t+1} - x^*\|^2 + \frac{a_{t+1}^2}{2\tau} \|v^{t+1} - u^*\|^2 - \frac{a_{t+1}^2 \gamma}{2 + 2a_{t+1} \gamma \mu_g} \|\mathbf{K}^*(v^{t+1} - u^*)\|^2 \\ & \leq \frac{1}{2\gamma} \|z^t - x^*\|^2 + \frac{a_t^2}{2\tau} \|v^t - u^*\|^2 - \frac{a_t^2 \gamma}{2 + 2a_t \gamma \mu_g} \|\mathbf{K}^*(v^t - u^*)\|^2 \\ & \quad - \left(\frac{1}{2\gamma} - \frac{L_f}{2} \right) \|z^{t+1} - z^t\|^2 - \frac{\tau}{2} \|\mathbf{K} z^{t+1} - \tilde{\nabla} h^*(v^{t+1})\|^2 \\ & \quad - a_{t+1}^2 ((D_f + D_g)(x^{t+1}) + D_{h^*}(u^{t+1})) + (a_{t+1}^2 - a_{t+1}) ((D_f + D_g)(x^t) + D_{h^*}(u^t)) \\ & \quad - \frac{a_{t+1} \mu_g}{2} \|z^{t+1} - x^*\|^2 - \frac{a_{t+1} \mu_{h^*}}{2} \|v^{t+1} - u^*\|^2 \\ & \quad - \frac{a_{t+1}}{2L_f} \|\nabla f(y^t) - \nabla f(x^*)\|^2 - \frac{a_{t+1}^2 - a_{t+1}}{2L_f} \|\nabla f(x^t) - \nabla f(y^t)\|^2. \end{aligned}$$

Finally, rearranging the terms and substituting the Lagrangian gap for the Bregman distances yields the inequality (21). \square

Note that, for any $t \geq 0$ and $u \in \mathcal{U}$, the condition $\gamma\tau\|\mathbf{K}\|^2 \leq 1 + a_t\gamma\mu_g$ guarantees that $\frac{a_{t+1}^2}{2\tau} \|u\|^2 - \frac{a_{t+1}^2 \gamma}{2 + 2a_{t+1} \gamma \mu_g} \|\mathbf{K}^* u\|^2 \geq 0$. This ensures that the corresponding terms in (21) remain nonnegative.

Remark 3.2. By applying Jensen's inequality to Step 8 in APAPC, we obtain, for every $t \geq 0$ and any dual solution u^* , the bound $\|u^{t+1} - u^*\|^2 \leq (1 - \frac{1}{a_{t+1}}) \|u^t - u^*\|^2 + \frac{1}{a_{t+1}} \|v^{t+1} - u^*\|^2$. Consequently, if we have $\|v^t - u^*\|^2 \rightarrow 0$ and $\sum_{t \geq 1} \frac{1}{a_t} = \infty$, then $\|u^t - u^*\|^2 \rightarrow 0$ as well [Polyak, 1987, Lemma 3].

3.2 Accelerated Convergence with h Smooth

In this section, we consider the case where $\mu_{h^*} > 0$. Equivalently, h is L_h -smooth with $L_h = 1/\mu_{h^*}$. In this regime, the dual problem (20) is strongly convex, ensuring that its solution u^* is unique.

As a consequence of Theorem 3.1, we obtain the following convergence guarantees.

Theorem 3.3 (Accelerated convergence of APAPC with $\mu_{h^*} > 0$). *Suppose that $\mu_{h^*} > 0$ and let (x^*, u^*) be a solution of (3). In APAPC, define the Lyapunov function*

$$\begin{aligned} \mathcal{E}_{x^*}^t & := \frac{1 + a_t \gamma \mu_g}{2\gamma} \|z^t - x^*\|^2 + \frac{a_t^2 + a_t \tau \mu_{h^*}}{2\tau} \|v^t - u^*\|^2 - \frac{a_t^2 \gamma}{2 + 2a_t \gamma \mu_g} \|\mathbf{K}^*(v^t - u^*)\|^2 \\ & \quad + a_t^2 \mathcal{G}_{x^*, u^*}(x^t, u^t), \quad \forall t \geq 0. \end{aligned}$$

Assume that $\gamma \leq \frac{1}{L_f}$, $a_0 > 0$, $\gamma\tau\|\mathbf{K}\|^2 \leq 1 + a_0\gamma\mu_g$, and $a_{t+1} \geq a_t$ for every $t \geq 0$. Then

$$\mathcal{E}_{x^*}^{t+1} \leq \max\left(\frac{1}{1 + a_t\gamma\mu_g}, \frac{a_{t+1}^2}{a_t^2 + a_t\tau\mu_{h^*}}, \frac{a_{t+1}^2 - a_{t+1}}{a_t^2}\right) \mathcal{E}_{x^*}^t, \quad \forall t \geq 0.$$

Consequently, if, in addition, $a_{t+1}^2 - a_{t+1} \leq a_t^2$ and $a_{t+1}^2 \leq a_t^2 + a_t\tau\mu_{h^*}$ for every $t \geq 0$ (note that since $a_1 \geq 1$ is required, we must initialize $a_0 \geq \frac{\sqrt{\tau^2\mu_{h^*}^2+4}-\tau\mu_{h^*}}{2}$), then

$$\mathcal{E}_{x^*}^t \leq \mathcal{E}_{x^*}^0, \quad \forall t \geq 0.$$

This ensures the Lagrangian gap is bounded by

$$\mathcal{G}_{x^*, u^*}(x^t, u^t) \leq \frac{\mathcal{E}_{x^*}^0}{a_t^2}, \quad \forall t \geq 0.$$

In particular, by choosing $a_0 := \frac{\sqrt{\tau^2\mu_{h^*}^2+4}-\tau\mu_{h^*}}{2}$, $a_1 = 1$, and

$$a_{t+1} := \min\left(\sqrt{a_t^2 + a_t\tau\mu_{h^*}}, \frac{1+\sqrt{1+4a_t^2}}{2}\right), \quad \forall t \geq 1,$$

which grows as $a_t = \Omega\left(\frac{t}{2} \min(1, \tau\mu_{h^*})\right)$, APAPC achieves an accelerated sublinear convergence rate of

$$\mathcal{G}_{x^*, u^*}(x^t, u^t) = \mathcal{O}\left(\frac{\mathcal{E}_{x^*}^0}{\min(1, \tau\mu_{h^*})^2 t^2}\right).$$

This implies $\|u^t - u^*\|^2 = \mathcal{O}(1/t^2)$ as well.

Corollary 3.4 (Accelerated linear convergence of APAPC with $\mu_g > 0$ and $\mu_{h^*} > 0$). Under the conditions of Theorem 3.3, suppose additionally that $L_f \geq \mu_g > 0$. Then the solution (x^*, u^*) of (3) is unique. Let $a_{\sharp} := \sqrt{\frac{L_f}{\mu_g}} \geq 1$. In APAPC, choose $\gamma = \min\left(\frac{1}{L_f}, \frac{1}{\|\mathbf{K}\|} \sqrt{\frac{\mu_{h^*}}{L_f}}\right)$, $\tau = \frac{1}{\gamma\|\mathbf{K}\|^2}$, $a_0 := \frac{\sqrt{\tau^2\mu_{h^*}^2+4}-\tau\mu_{h^*}}{2}$, $a_1 = 1$, and

$$a_{t+1} := \min\left(\sqrt{a_t^2 + a_t\tau\mu_{h^*}}, \frac{1+\sqrt{1+4a_t^2}}{2}, a_{\sharp}\right), \quad \forall t \geq 1.$$

Let $T \geq 1$ such that $a_T = a_{\sharp}$. Then APAPC achieves accelerated linear convergence:

$$\mathcal{E}_{x^*}^{t+1} \leq \max\left(\frac{1}{1 + \sqrt{\frac{\mu_g}{L_f}}}, \frac{1}{1 + \frac{\sqrt{\mu_g\mu_{h^*}}}{\|\mathbf{K}\|}}\right) \mathcal{E}_{x^*}^t, \quad \forall t \geq T.$$

Consequently, its iteration complexity (to reach an accuracy $\mathcal{E}_{x^*}^t \leq \epsilon$) is

$$\mathcal{O}\left(\left(\sqrt{\frac{L_f}{\mu_g}} + \frac{\|\mathbf{K}\|}{\sqrt{\mu_g\mu_{h^*}}}\right) \log\left(\frac{\mathcal{E}_{x^*}^0}{\epsilon}\right)\right). \quad (25)$$

Proof. We choose $\tau = \frac{1}{\gamma\|\mathbf{K}\|^2}$ and seek $a_\# \geq 1$ and $\gamma \leq \frac{1}{L_f}$ that minimize the maximum of the three terms: (i) $\frac{1}{1+a_\#\gamma\mu_g} \geq \frac{1}{1+a_\#\frac{\mu_g}{L_f}}$, (ii) $\frac{1}{1+a_\#\gamma\|\mathbf{K}\|^2} \geq \frac{1}{1+a_\#\frac{\mu_{h^*}}{L_f}}$, (iii) $1 - \frac{1}{a_\#} \leq \frac{1}{1+\frac{1}{a_\#}}$. By choosing $a_\# = \sqrt{\frac{L_f}{\mu_g}}$, (i) dominates (iii). Also, (i) dominates (ii) provided $\gamma \leq \frac{1}{\|\mathbf{K}\| \sqrt{\frac{\mu_{h^*}}{L_f}}}$. This leads to the stated choice of γ , and the contraction factor is $\frac{1}{1+a_\#\gamma\mu_g} = \max\left(\left(1 + \sqrt{\frac{\mu_g}{L_f}}\right)^{-1}, \left(1 + \frac{\sqrt{\mu_g\mu_{h^*}}}{\|\mathbf{K}\|}\right)^{-1}\right)$. \square

Remark 3.5. If $a_t \equiv 1$, **APAPC** reduces to **PAPC**. Under the conditions of Corollary 3.4, with $\gamma = \min\left(\frac{1}{L_f}, \frac{1}{\|\mathbf{K}\| \sqrt{\frac{\mu_{h^*}}{\mu_g}}}\right)$ and $\tau = \frac{1}{\gamma\|\mathbf{K}\|^2}$, the complexity of **PAPC** is

$$\mathcal{O}\left(\left(\frac{L_f}{\mu_g} + \frac{\|\mathbf{K}\|}{\sqrt{\mu_g\mu_{h^*}}}\right) \log\left(\frac{\mathcal{E}_{x^*}^0}{\epsilon}\right)\right).$$

The first term is not accelerated. In contrast, **APAPC** accelerates both the primal dependence on L_f/μ_g and the primal–dual coupling factor, as shown in (25).

Remark 3.6. When $h = 0$, letting $L_h \rightarrow 0$ (equivalently $\mu_{h^*} \rightarrow \infty$) in Theorem 3.3 and Corollary 3.4 correctly recovers the results established in Section 2 for minimizing $f + g$ using **APGD**.

Remark 3.7. When h is smooth, Problem (1) consists of minimizing $\tilde{f} + g$, where $\tilde{f} = f + h \circ \mathbf{K}$ is $(L_f + \|\mathbf{K}\|^2 L_h)$ -smooth. Such a problem can be solved directly by a primal method that evaluates the gradient $\nabla \tilde{f}(y^t) = \nabla f(y^t) + \mathbf{K}^* \nabla h(\mathbf{K}y^t)$. As established in Corollary 2.3, **APGD** applied to this formulation yields a complexity scaling as $\sqrt{\frac{L_f + \|\mathbf{K}\|^2 L_h}{\mu_g}}$, matching (25). However, **APAPC** offers practical advantages: because the two functions f and $h \circ \mathbf{K}$ are decoupled and processed via two distinct stepsizes (γ and τ), **APAPC** can achieve better hidden constants and faster convergence. Moreover, the proximal treatment of h enhances robustness: inaccuracies in L_h may destabilize explicit gradient steps in **APGD**, whereas the implicit update in **APAPC** remains stable under such misspecification.

3.3 Accelerated Convergence with \mathbf{K}^* Bounded Below

In this section, we consider the case where \mathbf{K}^* is bounded below; that is, we define

$$\lambda_{\min}(\mathbf{K}\mathbf{K}^*) := \inf_{u \in \mathcal{U}, u \neq 0} \frac{\|\mathbf{K}^*u\|^2}{\|u\|^2}, \quad (26)$$

and assume $\lambda_{\min}(\mathbf{K}\mathbf{K}^*) > 0$. If \mathcal{U} has finite dimension, $\lambda_{\min}(\mathbf{K}\mathbf{K}^*)$ is the smallest eigenvalue of $\mathbf{K}\mathbf{K}^*$. In this regime too, the dual problem (20) is strongly convex, ensuring that its solution u^* is unique.

Theorem 3.8 (Accelerated convergence of **APAPC** with $\lambda_{\min}(\mathbf{K}\mathbf{K}^*) > 0$). *Suppose that $\lambda_{\min}(\mathbf{K}\mathbf{K}^*) > 0$ and $\mu_g \leq \frac{L_f}{2}$, and let (x^*, u^*) be a solution of (3). In **APAPC**, define the Lyapunov function*

$$\begin{aligned} \mathcal{E}_{x^*}^t := & \frac{2 + a_t\gamma\mu_g}{4\gamma} \|z^t - x^*\|^2 + \frac{2a_t^2 + \min\left(\frac{1}{2}, \frac{1}{2a_t\gamma L_f}\right)\gamma\tau a_t^2 \lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4\tau} \|v^t - u^*\|^2 \\ & - \frac{a_t^2\gamma}{2 + 2a_t\gamma\mu_g} \|\mathbf{K}^*(v^t - u^*)\|^2 + a_t^2 \mathcal{G}_{x^*, u^*}(x^t, u^t), \quad \forall t \geq 0. \end{aligned}$$

Assume that $\gamma \leq \frac{1}{2L_f}$, $a_0 > 0$, $\gamma\tau\|\mathbf{K}\|^2 \leq 1 + a_0\gamma\mu_g$, and $a_{t+1} \geq a_t$ for every $t \geq 0$. Then

$$\mathcal{E}_{x^*}^{t+1} \leq \max \left(\frac{1}{1 + \frac{1}{2}a_t\gamma\mu_g}, \frac{a_{t+1}^2}{a_t^2 + \frac{1}{4}a_t^2\gamma\tau\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}, \frac{a_{t+1}^2}{a_t^2 + \frac{1}{4\gamma L_f}a_t\gamma\tau\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}, \frac{a_{t+1}^2 - a_t^2}{a_t^2} \right) \mathcal{E}_{x^*}^t,$$

for every $t \geq 0$. Consequently, choosing $\tau = \frac{\nu}{\gamma\|\mathbf{K}\|^2}$ for some $\nu \in (0, 1]$, $a_0 = a_1 = 1$, and

$$a_{t+1} := \min \left(a_t \sqrt{1 + \frac{\nu\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4\|\mathbf{K}\|^2}}, \sqrt{a_t^2 + a_t \frac{\nu\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4\gamma L_f \|\mathbf{K}\|^2}}, \frac{1 + \sqrt{1 + 4a_t^2}}{2} \right), \quad \forall t \geq 1,$$

which grows as $a_t = \Omega \left(\frac{t}{2} \min \left(1, \frac{\nu\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4\gamma L_f \|\mathbf{K}\|^2} \right) \right)$, we obtain

$$\mathcal{E}_{x^*}^t \leq \mathcal{E}_{x^*}^0, \quad \forall t \geq 0.$$

Therefore,

$$\mathcal{G}_{x^*, u^*}(x^t, u^t) \leq \frac{\mathcal{E}_{x^*}^0}{a_t^2} = \mathcal{O} \left(\frac{1}{t^2} \right), \quad \forall t \geq 0.$$

Moreover, if $\nu < 1$, then $\|v^t - u^*\|^2 = \mathcal{O}(1/t^2)$; if $\nu = 1$, then $\|v^t - u^*\|^2 = \mathcal{O}(1/t)$.

Proof. Let (x^*, u^*) be a solution of (3) and $t \geq 0$. Starting from (21), we exploit the negative term $-\frac{1}{2\gamma}\|z^{t+1} - z^t\|^2$ to bring up a negative factor $-\|v^{t+1} - u^*\|^2$. To this end, we use Young's inequality $\|b + c + d\|^2 \geq \|b\|^2/2 - 2\|c\|^2 - 2\|d\|^2$ for any vectors b, c, d . This yields

$$\begin{aligned} \frac{1}{2\gamma}\|z^{t+1} - z^t\|^2 &= \frac{\gamma a_{t+1}^2}{2} \left\| \mu_g(z^{t+1} - x^*) + \nabla f(y^t) - \nabla f(x^*) + \mathbf{K}^*(v^{t+1} - u^*) \right\|^2 \\ &\geq \frac{\gamma a_{t+1}^2}{4} \left\| \mathbf{K}^*(v^{t+1} - u^*) \right\|^2 - a_{t+1}^2 \gamma \mu_g^2 \|z^{t+1} - x^*\|^2 - a_{t+1}^2 \gamma \left\| \nabla f(y^t) - \nabla f(x^*) \right\|^2. \end{aligned}$$

Define $\delta_{t+1} := \min \left(\frac{1}{2}, \frac{1}{2a_{t+1}\gamma L_f} \right)$. Then

$$\begin{aligned} -\frac{1}{2\gamma}\|z^{t+1} - z^t\|^2 &= -\frac{1 - \delta_{t+1}}{2\gamma}\|z^{t+1} - z^t\|^2 - \frac{\delta_{t+1}}{2\gamma}\|z^{t+1} - z^t\|^2 \\ &\leq -\frac{1}{4\gamma}\|z^{t+1} - z^t\|^2 - \frac{\delta_{t+1}\gamma a_{t+1}^2 \lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4} \|v^{t+1} - u^*\|^2 + \frac{a_{t+1}\mu_g^2}{2L_f} \|z^{t+1} - x^*\|^2 \\ &\quad + \frac{a_{t+1}}{2L_f} \left\| \nabla f(y^t) - \nabla f(x^*) \right\|^2, \end{aligned}$$

where we used $\delta_{t+1} \leq \frac{1}{2}$, $\delta_{t+1} \leq \frac{1}{2a_{t+1}\gamma L_f}$, and

$$\left\| \mathbf{K}^*(v^{t+1} - u^*) \right\|^2 \geq \lambda_{\min}(\mathbf{K}\mathbf{K}^*) \|v^{t+1} - u^*\|^2. \quad (27)$$

Substituting this bound into (21) gives

$$\begin{aligned}
& \frac{1 + a_{t+1}\gamma\mu_g}{2\gamma} \|z^{t+1} - x^*\|^2 + \frac{a_{t+1}^2}{2\tau} \|v^{t+1} - u^*\|^2 - \frac{a_{t+1}^2\gamma}{2 + 2a_{t+1}\gamma\mu_g} \|\mathbf{K}^*(v^{t+1} - u^*)\|^2 \\
& + a_{t+1}^2 \mathcal{G}_{x^*, u^*}(x^{t+1}, u^{t+1}) \\
& \leq \frac{1}{2\gamma} \|z^t - x^*\|^2 + \frac{a_{t+1}^2}{2\tau} \|v^t - u^*\|^2 - \frac{a_{t+1}^2\gamma}{2 + 2a_{t+1}\gamma\mu_g} \|\mathbf{K}^*(v^t - u^*)\|^2 \\
& - \left(\frac{1}{4\gamma} - \frac{L_f}{2} \right) \|z^{t+1} - z^t\|^2 - \frac{\tau}{2} \|\mathbf{K}z^{t+1} - \tilde{\nabla}h^*(v^{t+1})\|^2 \\
& + (a_{t+1}^2 - a_{t+1})\mathcal{G}_{x^*, u^*}(x^t, u^t) - \frac{\delta_{t+1}\gamma a_{t+1}^2 \lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4} \|v^{t+1} - u^*\|^2 + \frac{a_{t+1}\mu_g^2}{2L_f} \|z^{t+1} - x^*\|^2.
\end{aligned}$$

Using the assumptions $\mu_g \leq \frac{L_f}{2}$, $\gamma \leq \frac{1}{2L_f}$ and $\gamma\tau\|\mathbf{K}\|^2 \leq 1 + a_{t+1}\gamma\mu_g$, this simplifies the inequality to

$$\begin{aligned}
& \frac{2 + a_{t+1}\gamma\mu_g}{4\gamma} \|z^{t+1} - x^*\|^2 + \frac{2a_{t+1}^2 + \delta_{t+1}\gamma\tau a_{t+1}^2 \lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4\tau} \|v^{t+1} - u^*\|^2 \\
& - \frac{a_{t+1}^2\gamma}{2 + 2a_{t+1}\gamma\mu_g} \|\mathbf{K}^*(v^{t+1} - u^*)\|^2 + a_{t+1}^2 \mathcal{G}_{x^*, u^*}(x^{t+1}, u^{t+1}) \\
& \leq \frac{2}{4\gamma} \|z^t - x^*\|^2 + \frac{2a_{t+1}^2}{4\tau} \|v^t - u^*\|^2 - \frac{a_{t+1}^2\gamma}{2 + 2a_{t+1}\gamma\mu_g} \|\mathbf{K}^*(v^t - u^*)\|^2 \\
& + (a_{t+1}^2 - a_{t+1})\mathcal{G}_{x^*, u^*}(x^t, u^t) - \frac{\tau}{2} \|\mathbf{K}z^{t+1} - \tilde{\nabla}h^*(v^{t+1})\|^2.
\end{aligned} \tag{28}$$

This yields the claimed Lyapunov inequality. Finally, $\frac{a_t^2}{2\tau}(1 - \nu) \|v^t - u^*\|^2 \leq \mathcal{E}_{x^*}^t$, where $\nu = \gamma\tau\|\mathbf{K}\|^2 \leq 1$. If $\nu < 1$, this implies $\|v^t - u^*\|^2 = \mathcal{O}(1/a_t^2) = \mathcal{O}(1/t^2)$; also, \mathbf{Q}_t^t (defined in (22)) is positive definite, so $\langle \cdot, \cdot \rangle_{\mathbf{Q}_t}$ is an inner product. If $\nu = 1$, since $a_t \rightarrow \infty$, for sufficiently large t we have $\frac{a_t \lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{8L_f} \|v^t - u^*\|^2 \leq \mathcal{E}_{x^*}^t$, which yields $\|v^t - u^*\|^2 = \mathcal{O}(1/t)$. \square

Corollary 3.9 (Accelerated linear convergence of APAPC with $\mu_g > 0$ and $\lambda_{\min}(\mathbf{K}\mathbf{K}^*) > 0$). *Under the conditions of Theorem 3.8, suppose additionally that $\mu_g > 0$. Then the solution (x^*, u^*) of*

(3) *is unique. In APAPC, choose $\gamma \in \left[\frac{1}{2L_f} \sqrt{\frac{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{\|\mathbf{K}\|^2}}, \frac{1}{2L_f} \right]$, set $\tau = \frac{1}{\gamma\|\mathbf{K}\|^2}$, $a_0 = a_1 = 1$, let*

$$a_{\#} := \max \left(\frac{1}{\gamma} \sqrt{\frac{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{2\mu_g L_f \|\mathbf{K}\|^2}}, 1 \right), \text{ and}$$

$$a_{t+1} := \min \left(a_t \sqrt{1 + \frac{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4\|\mathbf{K}\|^2}}, \sqrt{a_t^2 + a_t \frac{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4\gamma L_f \|\mathbf{K}\|^2}}, \frac{1 + \sqrt{1 + 4a_t^2}}{2}, a_{\#} \right), \quad \forall t \geq 1.$$

Let $T \geq 1$ such that $a_T = a_{\#}$. Then APAPC achieves accelerated linear convergence:

$$\mathcal{E}_{x^*}^{t+1} \leq \max \left(\frac{1}{1 + \sqrt{\frac{\mu_g \lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{8L_f \|\mathbf{K}\|^2}}}, \frac{1}{1 + \frac{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4\|\mathbf{K}\|^2}} \right) \mathcal{E}_{x^*}^t, \quad \forall t \geq T.$$

Consequently, its iteration complexity (to reach an accuracy $\mathcal{E}_{x^*}^t \leq \epsilon$) is

$$\mathcal{O} \left(\left(\sqrt{\frac{L_f \|\mathbf{K}\|^2}{\mu_g \lambda_{\min}(\mathbf{K}\mathbf{K}^*)}} + \frac{\|\mathbf{K}\|^2}{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)} \right) \log \left(\frac{\mathcal{E}_{x^*}^0}{\epsilon} \right) \right). \quad (29)$$

Proof. We choose $\tau = \frac{1}{\gamma \|\mathbf{K}\|^2}$ and seek $a_{\sharp} \geq 1$ and $\gamma \leq \frac{1}{2L_f}$ that minimize the maximum of the four terms: (i) $\frac{1}{1 + \frac{1}{2}a_{\sharp}\gamma\mu_g}$, (ii) $\frac{1}{1 + \frac{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4\gamma L_f a_{\sharp} \|\mathbf{K}\|^2}}$, (iii) $1 - \frac{1}{a_{\sharp}} \leq \frac{1}{1 + \frac{1}{a_{\sharp}}}$, (iv) $\frac{1}{1 + \frac{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4\|\mathbf{K}\|^2}}$. First, we rewrite (ii) as

$$\frac{1}{1 + \frac{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4\gamma L_f a_{\sharp} \|\mathbf{K}\|^2}} = \max \left(\frac{1}{1 + \sqrt{\frac{\mu_g \lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{8L_f \|\mathbf{K}\|^2}}}, \frac{1}{1 + \frac{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{4\gamma L_f \|\mathbf{K}\|^2}} \right).$$

Since $\gamma \leq \frac{1}{2L_f}$, the second term is dominated by (iv). Moreover, since $\frac{1}{a_{\sharp}} = \min \left(\gamma \sqrt{\frac{2\mu_g L_f \|\mathbf{K}\|^2}{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}}, 1 \right) \geq \sqrt{\frac{\mu_g}{2L_f}}$, (ii) dominates (iii). Finally, for (i),

since $a_{\sharp}\gamma \geq \sqrt{\frac{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{2\mu_g L_f \|\mathbf{K}\|^2}}$, it follows that $\frac{1}{1 + \frac{1}{2}a_{\sharp}\gamma\mu_g} \leq \frac{1}{1 + \frac{1}{2}\sqrt{\frac{\mu_g \lambda_{\min}(\mathbf{K}\mathbf{K}^*)}{2L_f \|\mathbf{K}\|^2}}}$, which is dominated by (ii).

This yields the stated contraction factor. \square

The complexity of **APAPC** in (29) improves upon that of **PAPC**, which, with appropriately chosen stepsizes, scales as $\frac{L_f}{\mu_g} + \frac{\|\mathbf{K}\|^2}{\lambda_{\min}(\mathbf{K}\mathbf{K}^*)}$.

Remark 3.10. Consider the case where $\mathbf{K} = \mathbf{I}$ and $\mu_g = 0$, which implies $\|\mathbf{K}\| = \lambda_{\min}(\mathbf{K}\mathbf{K}^*) = 1$. By setting $\tau = 1/\gamma$ in **APAPC** and invoking the Moreau identity, the dual variable v^t cancels out and the primal update collapses to

$$\begin{aligned} z^{t+1} &= z^t - a_{t+1}\gamma \nabla f(y^t) - a_{t+1}\gamma \text{prox}_{\frac{1}{a_{t+1}\gamma} h^*} \left(v^t + \frac{1}{a_{t+1}\gamma} (z^t - a_{t+1}\gamma \nabla f(y^t) - a_{t+1}\gamma v^t) \right) \\ &= \text{prox}_{a_{t+1}\gamma h} \left(z^t - a_{t+1}\gamma \nabla f(y^t) \right). \end{aligned}$$

Thus, **APAPC** reduces exactly to **APGD**, with h playing the role of g . Consequently, Theorem 3.8 and Corollary 3.9 consistently recover the accelerated complexity established in Section 2 for minimizing $f + h$ via **APGD**.

3.4 Accelerated Convergence for Linearly Constrained Problems

Let $b \in \text{ran}(\mathbf{K})$, the range of \mathbf{K} . In this section, we consider the linearly constrained problem

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad f(x) + \frac{\mu_g}{2} \|x\|^2 \quad \text{s.t. } \mathbf{K}x = b. \quad (30)$$

This is a special case of Problem (19) with $h(u) = (0 \text{ if } u = b, +\infty \text{ otherwise})$. Its conjugate is $h^*(u) = \langle u, b \rangle$. The associated dual problem is therefore

$$\underset{u \in \mathcal{U}}{\text{minimize}} \quad \left(f + \frac{\mu_g}{2} \|\cdot\|^2 \right)^* (-\mathbf{K}^*u) + \langle u, b \rangle. \quad (31)$$

We define $\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)$ as the lower bound of the spectrum of $\mathbf{K}\mathbf{K}^*$ excluding zero; that is,

$$\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*) := \inf_{u \in \text{ran}(\mathbf{K}), u \neq 0} \frac{\|\mathbf{K}^*u\|^2}{\|u\|^2},$$

and we assume $\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*) > 0$ (this implies that $\text{ran}(\mathbf{K})$ is closed). This condition is only a safeguard for the infinite-dimensional case: if \mathcal{U} has finite dimension, $\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)$ is the smallest nonzero eigenvalue of $\mathbf{K}\mathbf{K}^*$, so no additional assumption on \mathbf{K} is required. Note that $\lambda_{\min}(\mathbf{K}\mathbf{K}^*)$, defined in (26), may still be zero.

In **APAPC**, we assume that $u^0 = v^0 \in \text{ran}(\mathbf{K})$. Step (5) simplifies to

$$v^{t+1} := v^t + \frac{\tau}{a_{t+1}}(\mathbf{K}z^t - b).$$

Therefore, by induction, $u^t, v^t \in \text{ran}(\mathbf{K})$ for every $t \geq 0$. This key property allows us to reuse the analysis of Section 3.3, with the improvement that $\lambda_{\min}(\mathbf{K}\mathbf{K}^*)$ can be replaced by $\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)$ in the crucial inequality (27). Equivalently, the algorithm can be viewed as operating on the real Hilbert space $\tilde{\mathcal{U}} := \text{ran}(\mathbf{K})$; redefining \mathbf{K} as a mapping from \mathcal{X} to $\tilde{\mathcal{U}}$ leads precisely to the definition of $\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)$. This property is specific to the linearly constrained setting, where prox_{h^*} maps $\text{ran}(\mathbf{K})$ into itself.

Accordingly, the results of Section 3.3 extend directly, with $\lambda_{\min}(\mathbf{K}\mathbf{K}^*)$ replaced by $\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)$. Although the dual problem (31) is not necessarily strongly convex on \mathcal{U} , its restriction to $\text{ran}(\mathbf{K})$ is strongly convex. We therefore define u^* as the unique solution of (31) in $\text{ran}(\mathbf{K})$. The optimality conditions are

$$\begin{cases} 0 \in \nabla f(x^*) + \mu_g x^* + \mathbf{K}^* u^*, \\ u^* \in \text{ran}(\mathbf{K}), \\ \mathbf{K}x^* = b, \end{cases} \quad (32)$$

and we assume that a primal solution x^* exists. The Lagrangian gap satisfies $\mathcal{G}_{x^*, u^*}(x, u) = (f + g)(x) - (f + g)(x^*) + \langle \mathbf{K}x - b, u^* \rangle = (f + g)(x) - (f + g)(x^*) - \langle x - x^*, \nabla f(x^*) + \mu_g x^* \rangle$ for every $x \in \mathcal{X}$ and $u \in \text{ran}(\mathbf{K})$. Since this quantity is independent of u and x^* , we denote it by $\mathcal{G}(x)$ in this section. Consequently, the analysis involves only the variable v^t , and u^t can be ignored. We now state the main result.

Theorem 3.11 (Accelerated convergence of **APAPC** for solving (32) with $\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*) > 0$). *Suppose that $\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*) > 0$ and $\mu_g \leq \frac{L_f}{2}$, and let (x^*, u^*) be a solution of (32). In **APAPC**, define the Lyapunov function*

$$\begin{aligned} \mathcal{E}_{x^*}^t := & \frac{2 + a_t \gamma \mu_g}{4\gamma} \|z^t - x^*\|^2 + \frac{2a_t^2 + \min(\frac{1}{2}, \frac{1}{2a_t \gamma L_f}) \gamma \tau a_t^2 \lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}{4\tau} \|v^t - u^*\|^2 \\ & - \frac{a_t^2 \gamma}{2 + 2a_t \gamma \mu_g} \|\mathbf{K}^*(v^t - u^*)\|^2 + a_t^2 \mathcal{G}(x^t) + \frac{a_t \tau}{2} \|\mathbf{K}x^t - b\|^2, \quad \forall t \geq 0. \end{aligned}$$

Assume that $u^0 = v^0 \in \text{ran}(\mathbf{K})$, $\gamma \leq \frac{1}{2L_f}$, $a_0 > 0$, $\gamma \tau \|\mathbf{K}\|^2 \leq 1 + a_0 \gamma \mu_g$, and $a_{t+1} \geq a_t$ for every $t \geq 0$. Then

$$\mathcal{E}_{x^*}^{t+1} \leq \max \left(\frac{1}{1 + \frac{1}{2} a_t \gamma \mu_g}, \frac{a_{t+1}^2}{a_t^2 + \frac{1}{4} a_t^2 \gamma \tau \lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}, \frac{a_{t+1}^2}{a_t^2 + \frac{1}{4\gamma L_f} a_t \gamma \tau \lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}, \frac{a_{t+1}^2 - a_{t+1}}{a_t^2} \right) \mathcal{E}_{x^*}^t,$$

for every $t \geq 0$. Consequently, by choosing $\tau = \frac{\nu}{\gamma \|\mathbf{K}\|^2}$ for some $\nu \in (0, 1]$, $a_0 = a_1 = 1$, and

$$a_{t+1} := \min \left(a_t \sqrt{1 + \frac{\nu \lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}{4\|\mathbf{K}\|^2}}, \sqrt{a_t^2 + a_t \frac{\nu \lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}{4\gamma L_f \|\mathbf{K}\|^2}}, \frac{1 + \sqrt{1 + 4a_t^2}}{2} \right), \quad \forall t \geq 1,$$

which grows as $a_t = \Omega\left(\frac{t}{2} \min\left(1, \frac{\nu \lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}{4\gamma L_f \|\mathbf{K}\|^2}\right)\right)$, we obtain

$$\mathcal{E}_{x^*}^t \leq \mathcal{E}_{x^*}^0, \quad \forall t \geq 0.$$

Therefore, for every $t \geq 0$,

$$\mathcal{G}(x^t) \leq \frac{\mathcal{E}_{x^*}^0}{a_t^2} = \mathcal{O}\left(\frac{1}{t^2}\right), \quad \|\mathbf{K}x^t - b\|^2 \leq \frac{2\mathcal{E}_{x^*}^0}{a_t\tau} = \mathcal{O}\left(\frac{1}{t}\right).$$

Moreover, if $\nu < 1$, then $\|v^t - u^*\|^2 = \mathcal{O}(1/t^2)$; if $\nu = 1$, then $\|v^t - u^*\|^2 = \mathcal{O}(1/t)$.

Proof. The proof follows that of Theorem 3.8, with $\lambda_{\min}(\mathbf{K}\mathbf{K}^*)$ replaced by $\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)$. Let (x^*, u^*) be a solution of (32) and $t \geq 0$. Starting from (28) with $\tilde{\nabla} h^*(v^{t+1}) = b$, we obtain

$$\begin{aligned} & \frac{2 + a_{t+1}\gamma\mu_g}{4\gamma} \|z^{t+1} - x^*\|^2 + \frac{2a_{t+1}^2 + \delta_{t+1}\gamma\tau a_{t+1}^2 \lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}{4\tau} \|v^{t+1} - u^*\|^2 \\ & - \frac{a_{t+1}^2\gamma}{2 + 2a_{t+1}\gamma\mu_g} \|\mathbf{K}^*(v^{t+1} - u^*)\|^2 + a_{t+1}^2\mathcal{G}(x^{t+1}) \\ & \leq \frac{2}{4\gamma} \|z^t - x^*\|^2 + \frac{2a_{t+1}^2}{4\tau} \|v^t - u^*\|^2 - \frac{a_{t+1}^2\gamma}{2 + 2a_{t+1}\gamma\mu_g} \|\mathbf{K}^*(v^t - u^*)\|^2 \\ & + (a_{t+1}^2 - a_{t+1})\mathcal{G}(x^t) - \frac{\tau}{2} \|\mathbf{K}z^{t+1} - b\|^2. \end{aligned} \quad (33)$$

Using Jensen's inequality together with (8), we have $\|\mathbf{K}x^{t+1} - b\|^2 \leq \left(1 - \frac{1}{a_{t+1}}\right) \|\mathbf{K}x^t - b\|^2 + \frac{1}{a_{t+1}} \|\mathbf{K}z^{t+1} - b\|^2$. Therefore, $\|\mathbf{K}z^{t+1} - b\|^2 \geq a_{t+1} \|\mathbf{K}x^{t+1} - b\|^2 - (a_{t+1} - 1) \|\mathbf{K}x^t - b\|^2$. Hence,

$$\begin{aligned} & \frac{2 + a_{t+1}\gamma\mu_g}{4\gamma} \|z^{t+1} - x^*\|^2 + \frac{2a_{t+1}^2 + \delta_{t+1}\gamma\tau a_{t+1}^2 \lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}{4\tau} \|v^{t+1} - u^*\|^2 \\ & - \frac{a_{t+1}^2\gamma}{2 + 2a_{t+1}\gamma\mu_g} \|\mathbf{K}^*(v^{t+1} - u^*)\|^2 + a_{t+1}^2\mathcal{G}(x^{t+1}) + \frac{a_{t+1}\tau}{2} \|\mathbf{K}x^{t+1} - b\|^2 \\ & \leq \frac{2}{4\gamma} \|z^t - x^*\|^2 + \frac{2a_{t+1}^2}{4\tau} \|v^t - u^*\|^2 - \frac{a_{t+1}^2\gamma}{2 + 2a_{t+1}\gamma\mu_g} \|\mathbf{K}^*(v^t - u^*)\|^2 \\ & + (a_{t+1}^2 - a_{t+1})\mathcal{G}(x^t) + \frac{(a_{t+1} - 1)\tau}{2} \|\mathbf{K}x^t - b\|^2. \end{aligned}$$

Moreover, $\frac{a_{t+1}-1}{a_t} = \frac{a_{t+1}^2 - a_{t+1}}{a_t a_{t+1}} \leq \frac{a_{t+1}^2 - a_{t+1}}{a_t^2}$. This yields the claimed recursion. \square

Remark 3.12. Under the conditions of Theorem 3.11, we have $\|\mathbf{K}x^t - b\|^2 = \mathcal{O}(1/t)$, which is not accelerated. One way to obtain an accelerated rate is to modify the objective by introducing the penalty term $\tilde{f} = f + \eta \|\mathbf{K} \cdot - b\|^2$ for some $\eta > 0$. In this case, a rate $\mathcal{O}(1/t^2)$ for the Lagrangian gap directly yields the same rate for $\eta \|\mathbf{K}x^t - b\|^2$.

Corollary 3.13 (Accelerated linear convergence of APAPC on (32) with $\mu_g > 0$ and $\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*) > 0$). *Under the conditions of Theorem 3.11, suppose additionally that $\mu_g > 0$. Then the solution (x^*, u^*)*

of (32) is unique. In **APAPC**, choose $\gamma \in \left[\frac{1}{2L_f} \sqrt{\frac{\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}{\|\mathbf{K}\|^2}}, \frac{1}{2L_f} \right]$, $\tau = \frac{1}{\gamma\|\mathbf{K}\|^2}$, and $a_0 = a_1 = 1$.

Define $a_{\sharp} := \max\left(\frac{1}{\gamma} \sqrt{\frac{\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}{2\mu_g L_f \|\mathbf{K}\|^2}}, 1\right)$ and choose

$$a_{t+1} := \min\left(a_t \sqrt{1 + \frac{\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}{4\|\mathbf{K}\|^2}}, \sqrt{a_t^2 + a_t \frac{\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}{4\gamma L_f \|\mathbf{K}\|^2}}, \frac{1 + \sqrt{1 + 4a_t^2}}{2}, a_{\sharp}\right), \quad \forall t \geq 1.$$

Let $T \geq 1$ such that $a_T = a_{\sharp}$. Then **APAPC** achieves accelerated linear convergence:

$$\mathcal{E}_{x^*}^{t+1} \leq \max\left(\frac{1}{1 + \sqrt{\frac{\mu_g \lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}{8L_f \|\mathbf{K}\|^2}}}, \frac{1}{1 + \frac{\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}{4\|\mathbf{K}\|^2}}\right) \mathcal{E}_{x^*}^t, \quad \forall t \geq T.$$

Consequently, its iteration complexity (to reach an accuracy $\mathcal{E}_{x^*}^t \leq \epsilon$) is

$$\mathcal{O}\left(\left(\sqrt{\frac{L_f \|\mathbf{K}\|^2}{\mu_g \lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}} + \frac{\|\mathbf{K}\|^2}{\lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}\right) \log\left(\frac{\mathcal{E}_{x^*}^0}{\epsilon}\right)\right). \quad (34)$$

Remark 3.14. Chebyshev acceleration can further improve the complexity factor in (34) to $\sqrt{\frac{L_f \|\mathbf{K}\|^2}{\mu_g \lambda_{\min}^+(\mathbf{K}\mathbf{K}^*)}}$, which is optimal [Salim et al., 2022a]. This amounts to replacing the constraint $\mathbf{K}x = b$ in (30) by $\sqrt{P(\mathbf{K}^*\mathbf{K})}x = \sqrt{P(\mathbf{K}^*\mathbf{K})}x^*$ for a suitably chosen polynomial P , effectively applying $\mathbf{K}^*\mathbf{K}$ multiple times per iteration.

APAPC can be applied to decentralized optimization; see Salim et al. [2022a], Xu et al. [2020] for presentations and discussions of optimal algorithms in the strongly convex and general convex settings, respectively.

3.5 Point Convergence of **APAPC**

In regimes where both the primal and dual problems are strongly convex, the linear convergence of **APAPC** guarantees the strong convergence of the iterates to the unique saddle-point solution of (3); that is, $\|u^t - u^*\| \rightarrow 0$ and $\|x^t - x^*\| \rightarrow 0$. In this section, we therefore focus on the sublinear regime where $\mu_g = 0$ and only the dual problem is assumed to be strongly convex. Under these conditions, our preceding analysis guarantees that the dual iterates converge strongly to the unique dual solution u^* , but the convergence of the primal iterates remains to be established. To complete our analysis, we characterize the weak convergence of the primal iterates generated by **APAPC**.

Theorem 3.15 (Point Convergence of **APAPC**). *Assume that $g = 0$ and that the following hold in **APAPC**:*

- (i) $a_t \rightarrow \infty$ and $\sum_{t \geq 1} \frac{1}{a_t} = \infty$.
- (ii) The solution u^* of the dual problem (20) is unique, and $\|v^t - u^*\| \rightarrow 0$.
- (iii) The sequence $(z^t)_{t \geq 0}$ is bounded.
- (iv) $\mathcal{G}_{x^*, u^*}(x^t, u^t) \rightarrow 0$ for every saddle-point solution (x^*, u^*) to (3).
- (v) $\|\mathbf{K}z^t - \tilde{\nabla} h^*(v^t)\| \rightarrow 0$.

Then, the following hold:

- (a) $\|u^t - u^*\| \rightarrow 0$ and $\|y^t - x^t\| = \mathcal{O}(1/a_t)$.
- (b) For every $t \geq 0$, $\|\nabla f(x^t) + \mathbf{K}^*u^*\|^2 \leq 2L_f \mathcal{G}_{x^*, u^*}(x^t, u^t) \rightarrow 0$.
- (c) The sequence $(x^t)_{t \geq 0}$ is bounded, and all its weak cluster points are solutions to (19).
- (d) Every weak cluster point z^∞ of $(z^t)_{t \geq 0}$ satisfies $\mathbf{K}z^\infty \in \partial h^*(u^*)$.

In addition, assume the following:

- (vi) For every saddle-point solution (x^*, u^*) to (3), there exists a Lyapunov function $\mathcal{E}_{x^*}^t := \frac{1}{2\gamma} \|z^t - x^*\|^2 + a_t^2 \mathcal{G}_{x^*, u^*}(x^t, u^t) + \mathcal{R}^t$ for some dual term $\mathcal{R}^t \geq 0$ that does not depend on x^* . Moreover, $(\mathcal{E}_{x^*}^t)_{t \geq 0}$ is nonincreasing.
- (vii) For any pair of solutions (\hat{x}^*, u^*) and (\check{x}^*, u^*) to (3), we have $\mathbf{K}\hat{x}^* = \mathbf{K}\check{x}^*$.

Then, there exists a solution x^* to (19) such that $(x^t)_{t \geq 0}$ and $(y^t)_{t \geq 0}$ both converge weakly to x^* .

Proof. Assumptions (i) and (ii) imply that $\|u^t - u^*\| \rightarrow 0$, as established in Remark 3.2. Since $(z^t)_{t \geq 0}$ is assumed to be bounded, there exists $M > 0$ such that $\|z^t\| \leq M$ for every $t \geq 0$. Consequently, $(x^t)_{t \geq 0}$ is also bounded, because $x^0 = z^0$ and $\|x^{t+1}\| \leq \left(1 - \frac{1}{a_{t+1}}\right) \|x^t\| + \frac{1}{a_{t+1}} \|z^{t+1}\| \leq \max(\|x^t\|, M) \leq \dots \leq \max(\|x^0\|, M) = M$. Therefore, $(x^t)_{t \geq 0}$ has at least one weak cluster point. Furthermore, from the update in Step 3 of APAPC, we have $\|y^t - x^t\| = \frac{1}{a_{t+1}} \|x^t - z^t\|$. Since $(x^t - z^t)_{t \geq 0}$ is bounded, this yields $\|y^t - x^t\| = \mathcal{O}(1/a_t)$.

Let z^∞ be a weak cluster point of $(z^t)_{t \geq 0}$, meaning there exists a subsequence $(z^{t_j})_{j \geq 0}$ such that $z^{t_j} \rightharpoonup z^\infty$, where \rightharpoonup denotes weak convergence. Because \mathbf{K} is bounded, it is weakly continuous, which implies $\mathbf{K}z^{t_j} \rightharpoonup \mathbf{K}z^\infty$. By Assumption (v), we have $\|\mathbf{K}z^{t_j} - \tilde{\nabla} h^*(v^{t_j})\| \rightarrow 0$. Therefore, $\tilde{\nabla} h^*(v^{t_j}) \rightharpoonup \mathbf{K}z^\infty$. Moreover, by Assumption (ii), we know that $\|v^{t_j} - u^*\| \rightarrow 0$. The subdifferential ∂h^* is a maximally monotone operator, meaning its graph is strongly-weakly sequentially closed [Bauschke and Combettes, 2017, Proposition 20.38]. It follows that the limit point $(u^*, \mathbf{K}z^\infty)$ belongs to $\text{gra}(\partial h^*)$, where gra denotes the graph; that is, $\mathbf{K}z^\infty \in \partial h^*(u^*)$.

Now, let us characterize the weak cluster points of $(x^t)_{t \geq 0}$.

Step 1: The weak cluster points of $(x^t)_{t \geq 0}$ are primal solutions. Given any solution (x^*, u^*) to (3), which therefore satisfies $\nabla f(x^*) = -\mathbf{K}^*u^*$, we define the Bregman distance $D_f : x \in \mathcal{X} \mapsto f(x) - f(x^*) + \langle x - x^*, \mathbf{K}^*u^* \rangle$. For every $t \geq 0$, we have $\mathcal{G}_{x^*, u^*}(x^t, u^t) \geq D_f(x^t) \geq \frac{1}{2L_f} \|\nabla f(x^t) + \mathbf{K}^*u^*\|^2$ [Bubeck, 2015, Lemma 3.4]. Therefore, by Assumption (iv), $\|\nabla f(x^t) + \mathbf{K}^*u^*\| \rightarrow 0$. The gradient ∇f is a maximally monotone operator, meaning its graph is weakly-strongly sequentially closed [Bauschke and Combettes, 2017, Proposition 20.38]. Consequently, for every weak cluster point x^∞ of $(x^t)_{t \geq 0}$, the weak convergence of the subsequence combined with the strong convergence of the gradient implies that $\nabla f(x^\infty) = -\mathbf{K}^*u^*$. This fulfills the first optimality condition in (3).

Let x^∞ be a weak cluster point of $(x^t)_{t \geq 0}$. We now prove the second optimality condition, $\mathbf{K}x^\infty \in \partial h^*(u^*)$. Because ∂h^* is a maximally monotone operator, this inclusion is equivalent to showing that for every pair $(u, w) \in \text{gra}(\partial h^*)$, the inequality $\langle u^* - u, \mathbf{K}x^\infty - w \rangle \geq 0$ is satisfied [Bauschke and Combettes, 2017, Definition 20.20]. Let $(u, w) \in \text{gra}(\partial h^*)$. For every $k \geq 1$, monotonicity yields $\langle v^k - u, \tilde{\nabla} h^*(v^k) - w \rangle \geq 0$, which we can expand as

$$0 \leq \langle u^* - u, \mathbf{K}z^k - w \rangle + e_k, \quad \text{where } e_k := \langle v^k - u^*, \mathbf{K}z^k - w \rangle - \langle v^k - u, \mathbf{K}z^k - \tilde{\nabla} h^*(v^k) \rangle. \quad (35)$$

Since $(\mathbf{K}z^t - w)_{t \geq 0}$ is bounded, Assumptions (ii) and (v) imply that $e_k \rightarrow 0$. By unrolling Step 7 of APAPC, we observe that for every $t \geq 1$, x^t is a weighted average of all past estimates z^k ; that is, $x^t = \sum_{k=1}^t w_{k,t} z^k$, where the weights $w_{k,t} := \frac{1}{a_k} \prod_{j=k+1}^t \left(1 - \frac{1}{a_j}\right) \geq 0$ satisfy $\sum_{k=1}^t w_{k,t} = 1$.

Multiplying the inequality (35) by $w_{k,t}$ and summing over $k = 1, \dots, t$ yields

$$\langle u^* - u, \mathbb{K}x^t - w \rangle + \sum_{k=1}^t w_{k,t} e_k \geq 0. \quad (36)$$

By Assumption (i), $\sum_{j \geq 1} \frac{1}{a_j} = \infty$; thus, a standard result on infinite products guarantees that for every $k \geq 1$, $\lim_{t \rightarrow \infty} \prod_{j=k+1}^t \left(1 - \frac{1}{a_j}\right) = 0$, which in turn means $\lim_{t \rightarrow \infty} w_{k,t} = 0$. Consequently, the weights $w_{k,t}$ form a regular Toeplitz summation method, and the Toeplitz limit theorem ensures that $\sum_{k=1}^t w_{k,t} e_k \rightarrow 0$ [Knopp, 1951, Theorem 43.4]. Let $(t_j)_{j \geq 0}$ be the subsequence such that $x^{t_j} \rightharpoonup x^\infty$. Because \mathbb{K} is a bounded linear operator, it is weakly continuous, ensuring $\mathbb{K}x^{t_j} \rightharpoonup \mathbb{K}x^\infty$. Passing to the limit as $j \rightarrow \infty$ in (36), we obtain $\langle u^* - u, \mathbb{K}x^\infty - w \rangle \geq 0$. This confirms that $\mathbb{K}x^\infty \in \partial h^*(u^*)$. Therefore, (x^∞, u^*) is a valid solution to (3), meaning x^∞ is a primal solution to (19).

Step 2: Uniqueness of the weak cluster point. We now prove that under the additional assumptions (vi)–(vii), the weak cluster point of $(x^t)_{t \geq 0}$ is unique. Let \hat{x}^* and \check{x}^* be two weak cluster points of $(x^t)_{t \geq 0}$. By Step 1, both are primal solutions. To prove that $\hat{x}^* = \check{x}^*$, we define the sequences $\mathcal{A}^t := \|x^t - \hat{x}^*\|^2 - \|x^t - \check{x}^*\|^2$ and $\mathcal{B}^t := \|z^t - \hat{x}^*\|^2 - \|z^t - \check{x}^*\|^2$ for every $t \geq 0$. By Assumption (vi), there exist nonnegative, nonincreasing Lyapunov sequences $(\mathcal{E}_{\hat{x}^*}^t)_{t \geq 0}$ and $(\mathcal{E}_{\check{x}^*}^t)_{t \geq 0}$. Evaluating their difference, using (24) and applying Assumption (vii) yields

$$\begin{aligned} 2\gamma(\mathcal{E}_{\hat{x}^*}^t - \mathcal{E}_{\check{x}^*}^t) &= \|z^t - \hat{x}^*\|^2 - \|z^t - \check{x}^*\|^2 + 2\gamma a_t^2 (\mathcal{G}_{\hat{x}^*, u^*}(x^t, u^t) - \mathcal{G}_{\check{x}^*, u^*}(x^t, u^t)) \\ &= \|z^t - \hat{x}^*\|^2 - \|z^t - \check{x}^*\|^2 + 2\gamma a_t^2 \langle \mathbb{K}\check{x}^* - \mathbb{K}\hat{x}^*, u^t - u^* \rangle \\ &= \|z^t - \hat{x}^*\|^2 - \|z^t - \check{x}^*\|^2 = \mathcal{B}^t. \end{aligned}$$

Because $(\mathcal{E}_{\hat{x}^*}^t)_{t \geq 0}$ and $(\mathcal{E}_{\check{x}^*}^t)_{t \geq 0}$ are bounded below and nonincreasing, they both converge. Consequently, the sequence $(\mathcal{B}^t)_{t \geq 0}$ converges to some limit $\mathcal{B}^\infty \in \mathbb{R}$. Moreover, for every $t \geq 0$, we can expand the norms to obtain the algebraic identities:

$$\begin{aligned} \mathcal{B}^{t+1} &= -2\langle z^{t+1}, \hat{x}^* - \check{x}^* \rangle + \|\hat{x}^*\|^2 - \|\check{x}^*\|^2, \\ \mathcal{A}^{t+1} &= -2\langle x^{t+1}, \hat{x}^* - \check{x}^* \rangle + \|\hat{x}^*\|^2 - \|\check{x}^*\|^2, \\ \mathcal{A}^t &= -2\langle x^t, \hat{x}^* - \check{x}^* \rangle + \|\hat{x}^*\|^2 - \|\check{x}^*\|^2. \end{aligned}$$

From the momentum update in Step 7 of APAPC, we have $z^{t+1} = x^{t+1} + (a_{t+1} - 1)(x^{t+1} - x^t)$. Substituting this into the expressions above yields $\mathcal{B}^{t+1} = \mathcal{A}^{t+1} + (a_{t+1} - 1)(\mathcal{A}^{t+1} - \mathcal{A}^t)$. Because $a_t \rightarrow \infty$, there exists $T \geq 0$ such that $a_t > 1$ for every $t \geq T$, and we have $\sum_{t=T}^\infty \frac{1}{a_t - 1} \geq \sum_{t=T}^\infty \frac{1}{a_t} = \infty$. These conditions allow us to invoke Lemma A.4 of Boţ et al. [2025a], which guarantees that $\mathcal{A}^t \rightarrow \mathcal{B}^\infty$. Finally, let $(x^{t_j})_{j \geq 0}$ and $(x^{t'_j})_{j \geq 0}$ be two subsequences of $(x^t)_{t \geq 0}$ converging weakly to \hat{x}^* and \check{x}^* , respectively. Then $\mathcal{A}^{t_j} \rightarrow -\|\hat{x}^* - \check{x}^*\|^2$ and $\mathcal{A}^{t'_j} \rightarrow \|\check{x}^* - \hat{x}^*\|^2$. Because the entire sequence $(\mathcal{A}^t)_{t \geq 0}$ converges to a single limit, these subsequential limits must coincide. Hence, $\mathcal{B}^\infty = -\|\hat{x}^* - \check{x}^*\|^2 = \|\hat{x}^* - \check{x}^*\|^2$, which implies $\hat{x}^* = \check{x}^*$. It follows that $(x^t)_{t \geq 0}$ has exactly one weak cluster point x^* , and therefore converges weakly to x^* . Finally, because $\|y^t - x^t\|^2 \rightarrow 0$, the sequence $(y^t)_{t \geq 0}$ also converges weakly to x^* . \square

We note that regarding Assumption (vii) of Theorem 3.15, the relations $\mathbb{K}\hat{x}^* \in \partial h^*(u^*)$ and $\mathbb{K}\check{x}^* \in \partial h^*(u^*)$ hold. Therefore, a sufficient condition for $\mathbb{K}\hat{x}^* = \mathbb{K}\check{x}^*$ is simply that $\partial h^*(u^*)$ is

single-valued. This occurs, for instance, if h is strictly convex on \mathcal{X} . In particular, it holds in the linearly constrained setting of Section 3.4, where $\partial h^*(u^*) = \{b\}$. We formally state this convergence result as follows:

Theorem 3.16 (Point Convergence of APAPC for Linearly Constrained Problems). *Consider the linearly constrained minimization setting of Section 3.4, with $\lambda_{\min}^+(\mathbb{K}\mathbb{K}^*) > 0$ and $\mu_g = 0$. Let (x^*, u^*) be a solution of (32). In APAPC, define the Lyapunov function*

$$\begin{aligned} \mathcal{E}_{x^*}^t := & \frac{1}{2\gamma} \|z^t - x^*\|^2 + \frac{2a_t^2 + \min(\frac{1}{2}, \frac{1}{2a_t\gamma L_f})\gamma\tau a_t^2 \lambda_{\min}^+(\mathbb{K}\mathbb{K}^*)}{4\tau} \|v^t - u^*\|^2 \\ & - \frac{a_t^2\gamma}{2} \|\mathbb{K}^*(v^t - u^*)\|^2 + a_t^2 \mathcal{G}_{x^*, u^*}(x^t, u^t) + \frac{a_t\tau}{2} \|\mathbb{K}x^t - b\|^2, \quad \forall t \geq 0. \end{aligned}$$

Assume that $u^0 = v^0 \in \text{ran}(\mathbb{K})$, $\gamma \leq \frac{1}{2L_f}$, $\tau = \frac{\nu}{\gamma\|\mathbb{K}\|^2}$ for some $\nu \in (0, 1]$, $a_0 = a_1 = 1$, and

$$a_{t+1} := \min\left(a_t \sqrt{1 + \frac{\nu\lambda_{\min}^+(\mathbb{K}\mathbb{K}^*)}{4\|\mathbb{K}\|^2}}, \sqrt{a_t^2 + a_t \frac{\nu\lambda_{\min}^+(\mathbb{K}\mathbb{K}^*)}{4\gamma L_f \|\mathbb{K}\|^2}}, \frac{1 + \sqrt{1 + 4a_t^2}}{2}\right), \quad \forall t \geq 1.$$

Then, according to Theorem 3.11, the sequence $(\mathcal{E}_{x^*}^t)_{t \geq 0}$ is nonincreasing (which implies that $(z^t)_{t \geq 0}$ is bounded), $\mathcal{G}_{x^*, u^*}(x^t, u^t) = \mathcal{O}(1/t^2)$, $a_t = \Theta(t)$, and $\|v^t - u^*\|^2 = \mathcal{O}(1/t)$. Moreover, it follows from (33) in the proof of Theorem 3.11 that $\|\mathbb{K}z^t - \tilde{\nabla} h^*(v^t)\| = \|\mathbb{K}z^t - b\| \rightarrow 0$. Because $\mathbb{K}x^* = b$ for any primal solution x^* , Assumptions (i)–(vii) of Theorem 3.15 are all satisfied. Consequently, there exists a solution x^* to (30) such that $(x^t)_{t \geq 0}$ and $(y^t)_{t \geq 0}$ both converge weakly to x^* .

4 Conclusion

We introduced the Accelerated Proximal Alternating Predictor-Corrector algorithm (APAPC), successfully embedding Nesterov momentum into a fully split primal–dual method for structured convex optimization. By demonstrating that primal acceleration can be stabilized by exploiting the strong convexity of the dual problem, we established optimal $\mathcal{O}(1/t^2)$ sublinear convergence rates, as well as accelerated linear convergence, across several problem regimes. Furthermore, leveraging our unified Lyapunov framework, we characterized the weak convergence of the primal–dual iterates.

There are several promising avenues for future research. A natural extension is to generalize APAPC to solve Problem (1) with an arbitrary nonsmooth function g . However, while accelerated linear convergence is likely attainable when h is smooth (analogous to the results in Section 3.2), linear convergence will generally not hold for an arbitrary g in the settings of Sections 3.3 and 3.4 [Zhao, 2022]. Beyond deterministic optimization, future work will explore randomized variants of APAPC. The incorporation of stochastic gradient estimators [Zhao et al., 2019, Salim et al., 2022b, Zhao, 2022, Li et al., 2022, Ehrhardt et al., 2025] is of particular interest. Similarly, replacing exact proximal steps with stochastic or approximate evaluations [Chambolle et al., 2018, Condat and Richtárik, 2023, Condat et al., 2025, Combettes and Madariaga, 2025] represents a critical next step toward reducing the computational burden in large-scale inverse problems [Papoutsellis et al., 2025] and machine learning applications [Mishchenko et al., 2022, Condat et al., 2026].

Acknowledgements

This work was supported by funding from King Abdullah University of Science and Technology (KAUST):

- i) KAUST Baseline Research Scheme,
- ii) Center of Excellence for Generative AI (award no. 5940),
- iii) Competitive Research Grant (CRG) Program (award no. 6460).

References

- S. A. Alghunaim, E. K. Ryu, K. Yuan, and A. H. Sayed. Decentralized proximal gradient algorithms with linear convergence rates. *IEEE Trans. Automatic Control*, 66(6):2787–2794, June 2021.
- Z. Allen-Zhu and L. Orecchia. Linear coupling: An ultimate unification of gradient and mirror descent. preprint arXiv:1407.1537, 2014.
- H. Attouch and J. Peypouquet. The rate of convergence of Nesterov’s accelerated forward-backward method is actually faster than $1/k^2$. *SIAM J. Optim.*, 26(3):1824–1834, 2016.
- J.-F. Aujol, C. Dossal, H. Labarrière, and A. Rondepierre. FISTA restart using an automatic estimation of the growth parameter. *Journal of Optimization Theory and Applications*, 206:51, 2025.
- A. Auslender and M. Teboulle. Interior gradient and proximal methods for convex and conic optimization. *SIAM J. Optim.*, 16(3):697–725, 2006.
- F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Optimization with sparsity-inducing penalties. *Found. Trends Mach. Learn.*, 4(1):1–106, 2012.
- H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2nd edition, 2017.
- H. H. Bauschke and W. M. Moursi. Understanding FISTA’s weak convergence: A step-by-step introduction to the 2025 milestone. preprint arXiv:2601.15398, 2026.
- A. Beck. *First-Order Methods in Optimization*. MOS-SIAM Series on Optimization. SIAM, 2017.
- A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.
- R. I. Boţ, E. R. Csetnek, and C. Hendrich. Recent developments on primal–dual splitting methods with applications to convex minimization. In P. M. Pardalos and T. M. Rassias, editors, *Mathematics Without Boundaries: Surveys in Interdisciplinary Research*, pages 57–99. Springer New York, 2014.
- R. I. Boţ, E. Chenchene, E. R. Csetnek, and D. A. Hulett. Accelerating diagonal methods for bilevel optimization: Unified convergence via continuous-time dynamics. preprint arXiv:2505.14389, 2025a.

- R. I. Boş, J. Fadili, and D.-K. Nguyen. The iterates of Nesterov’s accelerated algorithm converge in the critical regimes. preprint arXiv:2510.22715, Oct. 2025b.
- L. Bottou, F. E. Curtis, and J. Nocedal. Optimization methods for large-scale machine learning. *SIAM review*, 60(2):223–311, 2018.
- S. Bubeck. Convex optimization: Algorithms and complexity. *Found. Trends Mach. Learn.*, 8(3–4): 231–357, 2015.
- V. Cevher, S. Becker, and M. Schmidt. Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics. *IEEE Signal Process. Mag.*, 31(5):32–43, 2014.
- A. Chambolle and C. Dossal. On the convergence of the iterates of the "Fast Iterative Shrinkage/Thresholding Algorithm". *J. Optim. Theory Appl.*, 166:968–982, 2015.
- A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40(1):120–145, May 2011.
- A. Chambolle and T. Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016a.
- A. Chambolle and T. Pock. On the ergodic convergence rates of a first-order primal–dual algorithm. *Math. Program.*, 159(1–2):253–287, Sept. 2016b.
- A. Chambolle, M. J. Ehrhardt, P. Richtárik, and C.-B. Schönlieb. Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications. *SIAM J. Optim.*, 28(4): 2783–2808, 2018.
- P. Chen, J. Huang, and X. Zhang. A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration. *Inverse Problems*, 29(2), 2013.
- Y. Chen, G. Lan, and Y. Ouyang. Optimal primal-dual methods for a class of saddle point problems. *SIAM J. Optim.*, 24:1779–1814, 2014.
- P. L. Combettes. The geometry of monotone operator splitting methods. *Acta Numerica*, 33: 487–632, 2024.
- P. L. Combettes and J. I. Madariaga. Almost-surely convergent randomly activated monotone operator splitting methods. *SIAM J. Imaging Sci.*, 18(4):2177–2205, 2025.
- P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer-Verlag, New York, 2010.
- P. L. Combettes and J.-C. Pesquet. Fixed point strategies in data science. *IEEE Trans. Signal Process.*, 69:3878–3905, 2021.
- P. L. Combettes, L. Condat, J.-C. Pesquet, and B. C. Vũ. A forward–backward view of some primal–dual optimization methods in image recovery. In *Proc. of IEEE ICIP*, Paris, France, Oct. 2014.

- L. Condat. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.*, 158(2):460–479, 2013.
- L. Condat and P. Richtárik. RandProx: Primal-dual optimization algorithms with randomized proximal updates. In *Proc. of Int. Conf. Learning Representations (ICLR)*, 2023.
- L. Condat, G. Malinovsky, and P. Richtárik. Distributed proximal splitting algorithms with rates and acceleration. *Frontiers in Signal Processing*, 1, Jan. 2022.
- L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi. Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists. *SIAM Review*, 65(2):375–435, 2023.
- L. Condat, E. Gasanov, and P. Richtárik. The stochastic multi-proximal method for nonsmooth optimization. preprint arXiv:2505.12409, 2025.
- L. Condat, I. Agarský, and P. Richtárik. CompressedScaffnew: The first theoretical double acceleration of communication from local training and compression in distributed optimization. *Optimization*, 2026.
- D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Set-Val. Var. Anal.*, 25:829–858, 2017.
- C. Dirren, M. Bianchi, P. D. Grontas, J. Lygeros, and F. Dörfler. Contractivity and linear convergence in bilinear saddle-point problems: An operator-theoretic approach. In *Proc. of Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, volume PMLR 258, 2025.
- D. Driggs, M. J. Ehrhardt, C.-B. Schönlieb, and J. Tang. Practical acceleration of the Condat–Vũ algorithm. *SIAM J. Imaging Sci.*, 17(4):2076–2109, 2024.
- Y. Drori, S. Sabach, and M. Teboulle. A simple algorithm for a class of nonsmooth convex concave saddle-point problems. *Oper. Res. Lett.*, 43(2):209–214, 2015.
- M. J. Ehrhardt, Z. Kereta, J. Liang, and J. Tang. A guide to stochastic optimisation for large-scale inverse problems. *Inverse Problems*, 41(5):053001, 2025.
- R. Glowinski, S. J. Osher, and W. Yin, editors. *Splitting Methods in Communication, Imaging, Science, and Engineering*. Springer International Publishing, 2016.
- U. Jang and E. K. Ryu. Point convergence of Nesterov’s accelerated gradient method: an AI-assisted proof. preprint arXiv:2510.23513, Oct. 2025.
- U. Jang, S. D. Gupta, and E. K. Ryu. Computer-assisted design of accelerated composite optimization methods: OptISTA. *Math. Program.*, 2025.
- K. Knopp. *Theory and Application of Infinite Series*. Blackie & Son, 2nd edition, 1951.
- D. Kovalev, A. Salim, and P. Richtárik. Optimal and practical algorithms for smooth and strongly convex decentralized optimization. In *Proc. of Conf. on Neural Information Processing Systems (NeurIPS)*, 2020.
- J. Lee, S. Yi, and E. K. Ryu. Convergence analyses of Davis–Yin splitting via scaled relative graphs. *SIAM J. Optim.*, 35(1):270–301, 2025.

- H. Li, Z. Lin, and Y. Fang. Variance reduced EXTRA and DIGing and their optimal acceleration for strongly convex decentralized optimization. *Journal of Machine Learning Research*, 23:1–41, 2022.
- I. Loris and C. Verhoeven. On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty. *Inverse Problems*, 27(12), 2011.
- K. Mishchenko, G. Malinovsky, S. Stich, and P. Richtárik. ProxSkip: Yes! Local Gradient Steps Provably Lead to Communication Acceleration! Finally! In *Proc. of the 39th International Conference on Machine Learning (ICML)*, July 2022.
- Y. Nesterov. *Introductory lectures on convex optimization: a basic course*. Kluwer Academic Publishers, 2004.
- Y. Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161, 2013.
- D. O’Connor and L. Vandenberghe. On the equivalence of the primal-dual hybrid gradient method and Douglas–Rachford splitting. *Math. Program.*, 79:85–108, 2020.
- D. P. Palomar and Y. C. Eldar, editors. *Convex Optimization in Signal Processing and Communications*. Cambridge University Press, 2009.
- E. Papoutsellis, Z. Kereta, and K. Papafitsoros. Why do we regularise in every iteration for imaging inverse problems? In *Proc. of International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 43–55. Springer Nature Switzerland, 2025.
- N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 3(1): 127–239, 2014.
- N. G. Polson, J. G. Scott, and B. T. Willard. Proximal algorithms in statistics and machine learning. *Statist. Sci.*, 30(4):559–581, 2015.
- B. T. Polyak. Gradient methods for the minimisation of functionals. *Ussr Computational Mathematics and Mathematical Physics*, 3:864–878, 1963.
- B. T. Polyak. *Introduction to Optimization*. Optimization Software, Inc., Publications Division, New York, NY, USA, 1987. Revised version of the 1987 original dated Nov. 2010 available on ResearchGate.
- A. Salim, L. Condat, D. Kovalev, and P. Richtárik. An optimal algorithm for strongly convex minimization under affine constraints. In *Proc. of Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, volume PMLR 151, pages 4482–4498, 2022a.
- A. Salim, L. Condat, K. Mishchenko, and P. Richtárik. Dualize, split, randomize: Toward fast nonsmooth optimization algorithms. *J. Optim. Theory Appl.*, July 2022b.
- S. Sra, S. Nowozin, and S. J. Wright. *Optimization for Machine Learning*. The MIT Press, 2011.
- G. Stathopoulos, H. Shukla, A. Szucs, Y. Pu, and C. N. Jones. Operator splitting methods in control. *Foundations and Trends in Systems and Control*, 3(3):249–362, 2016.

- A. B. Taylor, J. M. Hendrickx, and F. Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Math. Program.*, 161(1–2):307–345, 2017.
- A. B. Taylor, J. M. Hendrickx, and F. Glineur. Exact worst-case convergence rates of the proximal gradient method for composite convex minimization. *J. Optim. Theory Appl.*, 178:455–476, 2018.
- P. Tseng. On accelerated proximal gradient methods for convex–concave optimization. preprint, 2008.
- K. Ushiyama. A $\sqrt{2}$ -accelerated FISTA for composite strongly convex. preprint arXiv:2509.09295v2, Sept. 2025.
- B. C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3):667–681, Apr. 2013.
- Y. Wu, Y. Zhang, L. Liu, and Y. Ouyang. A note on the gradient-evaluation sequence in accelerated gradient methods. preprint arXiv:2603.06937, 2026.
- J. Xu, Y. Tian, Y. Sun, and G. Scutari. Accelerated primal–dual algorithm for distributed smooth convex optimization over networks. In *Proc. of International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume PMLR 108, 2020.
- M. Yan. A new primal-dual algorithm for minimizing the sum of three functions with a linear operator. *J. Sci. Comput.*, 76(3):1698–1717, Sept. 2018.
- R. Zhao. Accelerated stochastic algorithms for convex-concave saddle-point problems. *Math. Oper. Res.*, 47:1443–1473, 2022.
- R. Zhao, W. B. Haskell, and V. Y. F. Tan. An optimal algorithm for stochastic three-composite optimization. In *Proc. of International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume PMLR 89, 2019.
- Y.-N. Zhu. Accelerated primal dual fixed point algorithm. preprint arXiv:2511.00385, 2025.